

Final Technical Report • September 1995

IR MATERIALS PRODUCIBILITY

M.A. Berding, Sr. Research Physicist
A. Sher, Program Director
Physical Electronics Laboratory

SRI Project 3820

Prepared for:

Contracting Officer's Technical Representative
Advanced Research Projects Agency
Microelectronics Technology Office (MTO)
3701 N. Fairfax Drive
Arlington, VA 22203-1714

Attn: Mr. Raymond Balcerak

ARPA Order No. 8557; Program Code Nos. 2H20, 2D10

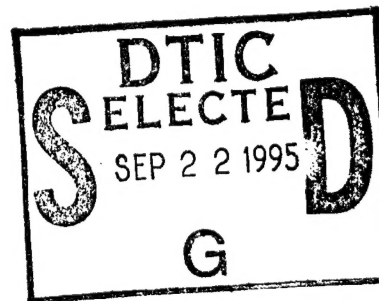
Contract MDA972-92-C-0053

Covering the period: 12 August 1992 through 11 September 1995

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or the U.S. Government.

APPROVED FOR PUBLIC RELEASE
DISTRIBUTION UNLIMITED

DTIC QUALITY INSPECTED 6



19950919 237

IR MATERIALS PRODUCIBILITY

M.A. Berding, Sr. Research Physicist
A. Sher, Program Director
Physical Electronics Laboratory

SRI Project 3820

Prepared for:

Contracting Officer's Technical Representative
Advanced Research Projects Agency
Microelectronics Technology Office (MTO)
3701 N. Fairfax Drive
Arlington, VA 22203-1714

Attn: Mr. Raymond Balcerak

Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	

ARPA Order No. 8557; Program Code Nos. 2H20, 2D10

Contract MDA972-92-C-0053

Covering the period: 12 August 1992 through 11 September 1995

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or the U.S. Government.

APPROVED FOR PUBLIC RELEASE
DISTRIBUTION UNLIMITED

Approved:

Eric Pearson, Director
Physical Electronics Laboratory

Donald L. Nielson, Vice President

SRI International Computing and Engineering Sciences Division
333 Ravenswood Avenue • Menlo Park, CA 94025-3493 • (415) 326-6200 • FAX: (415) 326-5512 • Telex: 334486



September 12, 1995

Mr. Raymond Balcerak, COTR
Advanced Research Projects Agency
Microelectronics Technology Office (MTO)
3701 North Fairfax Drive
Arlington, Virginia 22203-1714

Reference: Contract No. MDA972-92-C-0053
ARPA Order No. 8557
(SRI Project ECU-3820)

Dear Mr. Balcerak:

SRI International is pleased to submit the Final Technical Report entitled "IR Materials Producibility" for the referenced contract. This report is submitted in accordance with CDRL A004 and completes the requirements of the contract.

If you have any questions, please feel free to contact me at (415) 859-4328; my fax number is (415) 859-6171.

Sincerely,

Barbara E. Camph
Group Manager, Engineering Contracts

BEC:dl

Enclosure

cc: ARPA/OASB/Library (1 copy)
Defense Technical Information Center (2 copies)
Ms. Elaine Ely, ARPA/CMO (letter only)
Ms. Ligia P. Hazen, ACO/DCMAO-SF (letter only)

SRI International

333 Ravenswood Ave. • Menlo Park, CA 94025 • (415) 326-6200 • TWX: 910-373-2046 • Telex: 334486 • Facsimile: (415) 326-5512

- Auto-compensation by isolated native point defects is not responsible for the p -type doping limitations in ZnSe-based materials.
- Excess generation of Frenkel defects at stacking faults is responsible for the operational degradation of ZnSe-based LEDs and lasers.

CONTENTS

1	NATIVE DEFECTS IN HgCdTe	1
1.1	POINT DEFECTS	1
1.2	TELLURIUM-ANTISITE MERCURY-VACANCY PAIRS	3
1.3	LINE DEFECTS: DISLOCATIONS	4
2	NATIVE POINT DEFECTS IN CdTe	4
3	ELECTRONIC AND TRANSPORT PROPERTIES OF BULK HgCdTe	5
4	GRADIENT CORRECTIONS TO LDA	5
5	NATIVE POINT DEFECTS AND OPERATIONAL DEGRADATION IN ZnSe OPTICAL DEVICES	7
6	LIST OF PUBLICATIONS	9

1. NATIVE DEFECTS IN HgCdTe

1.1 POINT DEFECTS

We have developed an *ab initio* method to predict absolute densities of native point defects in semiconductors. At the heart of the calculation is a first-principles method used to predict electronic contribution to the native point defect free energies. This first-principles method is based on the local density approximation (LDA) to the many-body Schrödinger equation. We have also included the gradient corrections (GC) to the LDA, which we have shown are essential to obtaining quantitative agreement with the experimental results (see Section 4 below).

We use a more classical approach based on a valence force field model (VFF) to describe the vibrational contributions to the defect free energies. The VFF model has two independent elastic constants corresponding to bond length and bond angle distortions. We have demonstrated that in zincblende solids the three unique elastic constants permitted by symmetry (e.g. C_{11} , C_{12} , and C_{44}) are functionally related; thus, to a good approximation, only two independent constants exist. The magnitude of the effective constants that appear in the VFF can be expressed as linear combinations of C_{11} , C_{12} , and C_{44} , predicted from our first-principles theory. The agreement between the VFF parameters determined from fits to experiments and those calculated is excellent, differing by $\sim 10\%$ or less. Our procedure therefore has a firm theoretical base.

We have used a quasichemical analysis to deduce the defect concentrations, including full Fermi-Dirac statistics for the electron-hole populations. Eight classes of defects are included (simultaneously) in the analysis of $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$:

- V_{Hg} , the mercury vacancy
- V_{Te} , the tellurium vacancy
- Hg_{Te} , the mercury antisite
- Te_{Hg} , the tellurium antisite
- $\text{Hg}_{I_{\text{Hg}}}$, the mercury interstitial at a tetrahedral interstitial site, with mercury as the nearest neighbors
- $\text{Hg}_{I_{\text{Te}}}$, the mercury interstitial at a tetrahedral interstitial site, with tellurium as the nearest neighbors

- Te_{Hg} , the tellurium interstitial at a tetrahedral interstitial site, with mercury as the nearest neighbors
- Te_{Te} , the tellurium interstitial at a tetrahedral interstitial site, with tellurium as the nearest neighbors.

Although we have not included the cadmium antisite and cadmium interstitial (Cd_{Te} and Cd_I respectively) in the analysis, they are expected to be minority species and not affect device characteristics. The cadmium and mercury vacancies are indistinguishable and are both included in the analysis.

Our findings are summarized as follows:

- Mercury vacancies are the dominant defects at all temperatures and pressures, in agreement with experiment.
- Mercury vacancies are double acceptors, in agreement with experiment.
- Predicted carrier concentrations in nominally undoped material are in quantitative agreement with experiment.
- The tellurium antisite, a donor, is found to be an important, relatively plentiful, defect; moreover it
 - is a relatively slow diffuser (see Section 1.2 below),
 - may be the residual donor in liquid phase epitaxial (LPE) grown material,
 - may be a Shockley–Read recombination center, and
 - can be reduced in densities through multiple step anneals.
- The tellurium antisite mercury vacancy complex is well bound and, depending on processing details, can be present in concentrations high enough to affect devices.

We have proposed a series of experiments to test our predictions on the tellurium antisite. We have also examined how the defect densities are modified in the presence of second-phase tellurium. For additional details of this work, see the papers included as Appendices A, B, and C.

1.2 TELLURIUM-ANTISITE MERCURY-VACANCY PAIRS

We have predicted that the mercury vacancy and the tellurium antisite are, respectively, the first and second most plentiful defects in $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$. If equilibrium is truly achieved during the low-temperature ($\sim 250^\circ\text{C}$) anneals to which LWIR HgCdTe is subjected during processing, the equilibrium tellurium antisite densities obtained will be insignificant and will have no impact on device performance nor on subsequent processing. In fact, equilibrium densities of tellurium antisites may not be achieved during low-temperature anneals because of the antisite's low diffusion rates. While the diffusion mechanism of the mercury vacancy is simple (via the migration of an adjacent second-neighbor cation into the vacancy), the diffusion of the tellurium antisite will necessarily be more complex and involve another defect (for example, a cation vacancy) or a consorted exchange mechanism (whereby the tellurium antisite and an adjacent second-neighbor cation exchange positions).

Because of the more complex mechanism, we expect that the diffusion rates of tellurium antisites may be low. For liquid phase epitaxial (LPE) material, tellurium antisite densities may be frozen in at the high growth temperatures. The frozen-in tellurium antisite densities should be different for LPE material grown from the mercury and tellurium melt, primarily because they correspond to opposite sides of the existence region, but also because the two growths are done at slightly different temperatures.

We have considered the various means by which the tellurium antisite might diffuse and have determined that the mechanism involving the mercury vacancy is dominant. In this diffusion mechanism, motion of a tellurium antisite occurs by its migrating into an adjacent (second-neighbor) cation vacancy. The diffusion of the tellurium antisite will be proportional to

$$D_{\text{TeHg}} \sim [(V_{\text{Hg}} - \text{TeHg})_{\text{pair}}]. \quad (1)$$

Our calculations predict a very large binding energy of nearly 2 eV for this defect complex. We predict a substantial density of $(V_{\text{Hg}} - \text{TeHg})_{\text{pair}}$ present in equilibrated material. The pair is an *additional* source of tellurium antisites and mercury vacancies in the material. The density of the pairs will differ in LPE material grown from the mercury- and tellurium-rich melts, and will account for different rates of tellurium antisite diffusion of the two types of LPE material. We have proposed means by which

to manipulate the tellurium antisite densities via controlled cooling from the growth temperatures and through subsequent multiple step anneals. We have examined how tellurium antisites might contribute to precipitation of second-phase tellurium and have proposed a mechanism by which tellurium antisite diffusion is enhanced in a direction towards a tellurium precipitate. For additional discussion and details, see the paper that is included as Appendix A.

1.3 LINE DEFECTS: DISLOCATIONS

We have calculated the impact on junction performance of electric and strain fields associated with a 60° dislocation in HgCdTe. We find that the fields associated with core charges alone are not likely to account for the observed degradation of junction performance, except perhaps at very low dislocation densities and at cryogenic temperatures. The longer range piezoelectronic potentials, arising from strain fields, are likely to have a larger effect, especially on R_0A , through the tunneling leakage current. We propose that the nonlinear dependence of junction leakage current on junction potential is responsible for the observed nonlinearity of performance degradation with dislocation density as measured by etch pit density. Further details can be found in Appendices D and E.

2. NATIVE POINT DEFECTS IN CdTe

CdTe is important both as a substrate (along with $\text{Cd}_{1-x}\text{Zn}_x\text{Te}$ ¹) and passivating material for epitaxial layers of LWIR HgCdTe, making it an important material to successful manufacturing of HgCdTe FPAs. Native point defects are of interest in CdTe in that they relate to its stoichiometry (which in turn has been shown to affect the minority carrier lifetimes in HgCdTe devices), the formation and annihilation of tellurium precipitates, and the self- and inter-diffusion coefficients that affect material stability during growth, during subsequent processing, and over the device's lifetime.

Using our first-principles methods, we have predicted the density of native point defects in CdTe as a function of temperature and cadmium partial pressure. We predict that, unlike $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$, CdTe solid can exist in both cation- and anion-rich forms. We find that the cadmium interstitial is the dominant defect for high cadmium pressures, and the cadmium vacancy is most important at mid and low

¹Because of the low zinc concentrations, our results should be applicable to $x = 0.04$ $\text{Cd}_{1-x}\text{Zn}_x\text{Te}$, which is the more recent substrate of choice for LWIR $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ devices.

cadmium pressures. For further details, see Appendix A.

3. ELECTRONIC AND TRANSPORT PROPERTIES OF BULK HgCdTe

Most electronic property calculations on HgCdTe are done for zero temperature. We have included the electron-phonon interactions and predicted the temperature-dependent band gap in HgCdTe. These calculations are done using a semi-empirical hybrid pseudopotential tight-binding (HPTB) method. Our theory agrees with experiment to better than 10%. We have also used this theory to deduce the temperature variations of the valence band offsets and the electron effective masses of various semiconductor systems. The properties of $\text{In}_{0.33}\text{Tl}_{0.67}\text{P}$ with a band gap in the LWIR spectrum were also calculated and compared to those of $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$. Further details are included in Appendices F and G.

We have used our temperature-dependent HPTB method to predict the absorption edge and its temperature dependence. Calculations include the effects of the Urbach absorption tail, which is found experimentally to be more complex in HgCdTe than the simple exponential form usually assumed. The non-parabolicity of the conduction band edge is essential to the interpretation of the experimental results. Additional information is included in Appendix H.

Transport-related properties, including electron mobility, Hall coefficient, and the fundamental energy gap have been calculated using our semi-empirical HPTB band structures, and including full Fermi-Dirac statistics in the solution of the Boltzmann transport equation. Our results differ substantially from those obtained using $\mathbf{k}\cdot\mathbf{p}$ band structures. We have used our results to reinterpret absorption data and to extract the temperature variation of the band gap. Our calculated electron mobilities are in excellent agreement with experiment. For additional details, see Appendices G, I, J, and K.

4. GRADIENT CORRECTIONS TO LDA

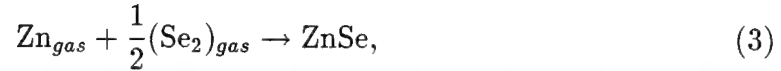
An essential ingredient to a quantitative prediction of native point defect densities in solids is an accurate evaluation of the energy to create the defect in the solid. It is often convenient to use the vapor phase as the reference phase, establishing the chemical potential of the constituents in the solids. In HgCdTe we have found the vapor phase to be the most natural reference phase because the mercury pressure

(P_{Hg}) is often used as a control parameter during processing. In ZnSe the vapor phase is also a convenient reference because the pressures over the ZnSe solid at high temperatures have been measured and because the beam fluxes in molecular beam epitaxy (MBE) can be related to the partial pressures of zinc and selenium.

Unfortunately, a significant overbinding is found for all of the semiconductors when the LDA is used. We have found (Appendix L) that the addition of gradient corrections to the LDA dramatically improves of the calculations, in particular when the energy of the solid is referenced to the vapor phase as is done in determining the cohesive energy, for example:



or



or when determining the energy to form a native defect where the vapor phase of one or more of the constituents is used as a reference phase, for example for a mercury vacancy in HgCdTe via the reaction



or a zinc antisite in ZnSe via the reaction



We have performed extensive tests of the gradient corrections on a wide range of solids, examining the changes made (relative to the LDA) in the predicted lattice constants, cohesive energy, and various elastic properties. These studies demonstrated systematic improvements in the predicted properties that are relevant to accurate prediction of the point defect formation energies. The gradient corrections were incorporated into the calculations of the native point defect densities of HgCdTe, CdTe, and ZnSe presented in the report. The results of our examination of 58 elemental and compound solids are being written for publication; a draft of this paper is included as Appendix L.

5. NATIVE POINT DEFECTS AND OPERATIONAL DEGRADATION IN ZnSe OPTICAL DEVICES

We have predicted the absolute defect concentrations of eight native point defects in ZnSe as a function of temperature and zinc partial pressure. The eight defects included in the analysis are:

- V_{Zn} , the zinc vacancy
- V_{Se} , the selenium vacancy
- Zn_{Se} , the zinc antisite
- Se_{Zn} , the selenium antisite
- $\text{Zn}_{I_{\text{Zn}}}$, the zinc interstitial at a tetrahedral interstitial site, with zinc as the nearest neighbors
- $\text{Zn}_{I_{\text{Se}}}$, the zinc interstitial at a tetrahedral interstitial site, with selenium as the nearest neighbors
- $\text{Se}_{I_{\text{Zn}}}$, the selenium interstitial at a tetrahedral interstitial site, with zinc as the nearest neighbors
- $\text{Se}_{I_{\text{Se}}}$, the selenium interstitial at a tetrahedral interstitial site, with selenium as the nearest neighbors.

We have used our state-of-the-art first-principles code based on the LDA to calculate the total electronic contribution to the defect formation free energy. We have also included the gradient corrections (GC) to the LDA, which we have shown are essential to obtaining quantitative agreement with the experimental results (see Section 4.) A classical approach based on a valence force field model is used to describe the vibrational contributions to the defect free energies. A quasichemical analysis was used to deduce the defect concentrations, including full Fermi-Dirac statistics for the electron-hole populations. Our results demonstrated that isolated native point defects are not responsible for the limitation in the p -type doping limits experimentally observed in ZnSe.

To investigate the operational optical degradation of ZnSe-based devices, we have looked at the total energy to form a zinc Frenkel defect. The zinc Frenkel defect is

a zinc-interstitial zinc-vacancy pair. The Frenkel defect is a particularly interesting defect (really a defect-generating process) in that it can form in the bulk of the solid, without the need for an external source or sink for atoms (as was needed for the reactions enumerated at the beginning of this section). We have found that the energy to form a Frenkel defect in the bulk zincblende crystal is slightly larger than the electron-hole recombination energy. We have reexamined the formation energy at a stacking fault and found a reduction of 0.3 eV in the energy to form the Frenkel defect. The reduction in energy at the stacking fault is large enough so that a Frenkel defect can be formed in a single nonradiative recombination event at a stacking fault. Our results also indicate that the zinc interstitial (part of a Frenkel defect) has a mid-gap state, both when trapped at a stacking fault and when in the bulk material, and should be an effective non-radiative recombination site.

Based on our findings, we propose a degradation mechanism in which excessive Frenkel defect pairs are generated at stacking faults present in a quantum well structure via the energy available through nonradiative recombination events. The zinc interstitials generated in these events serve as additional nonradiative recombination centers, thus amplifying the Frenkel defect generation rate. The thermal energy made available from nonradiative recombination events at the Frenkel-defect-dressed stacking faults produce temperature gradients that could also contribute to dislocation motion and multiplication.

This model may also provide the basis of an explanation for the so-called dark line defects. Some of the zinc interstitials that are formed at the stacking fault should be trapped there in potential wells, but others may diffuse away. In the presence of all of the dislocation motion and multiplication, one may have an ordering of the interstitials along the $\langle 100 \rangle$ directions. Although we have not completed this portion of our modeling, certainly our model of Frenkel defect generation at the stacking fault explanations a source of very mobile point defects that may be correlated with these dark line defects. A draft paper on our work is included as Appendix M.

6. LIST OF PUBLICATIONS

1. Defect modeling studies in HgCdTe and CdTe; M. A. Berding, A. Sher, and M. van Schilfgaarde, *J. Electron. Mater.* **24**, 1129 (1995).
2. First-principles calculation of native defect densities in $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$; M. A. Berding, M. van Schilfgaarde, and A. Sher, *Phys. Rev. B* **50**, 1519 (1994).
3. $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$ native defects: densities and dopant properties; M. A. Berding, M. van Schilfgaarde, and A. Sher, *J. Electron. Mater.* **22**, 1005 (1993).
4. How dislocations affect transport; A. T. Paxton, A. Sher, M. Berding, M. van Schilfgaarde, and M. W. Muller, *J. Electron. Mater.* **24**, 525 (1995).
5. Dislocations and junction performance in MCT; M. W. Muller A. T. Paxton, A. Sher, M. Berding, and M. van Schilfgaarde, Proceedings of the IRIS Materials Meeting, August 1993.
6. Temperature dependence of band gaps in HgCdTe and other semiconductors; S. Krishnamurthy, A.-B. Chen, A. Sher, and M. van Schilfgaarde, *J. Electron. Mater.* (in press).
7. InTlP: A superior IRFPA alloy; A. Sher, M. van Schilfgaarde, A.-B. Chen, and S. Krishnamurthy, Proceedings of the IRIS Materials Meeting, August 1994.
8. A study of the HgCdTe absorption edge at various temperatures; V. Ariel, V. Garber, G. Bahir, S. Krishnamurthy, and A. Sherm (Abstract submitted to 1995 MCT Workshop, to be published in *J. Electron. Mater.*).
9. Transport studies in narrow-gap semiconductors revisited; S. Krishnamurthy and A. Sher, *J. Electron. Mater.* **24**, 641 (1995).

10. Transport studies in narrow-gap semiconductors revisited; S. Krishnamurthy and A. Sher, Proceedings of the IRIS Materials Meeting, August 1993.
11. Electron mobility in $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ alloy; S. Krishnamurthy and A. Sher, *J. Appl. Phys.* **75**, 7904 (1994),
12. Systematic study of gradient corrections in a wide range of elemental and compound materials; M. A. Berding and M. van Schilfgaarde (In preparation for submission to *Phys. Rev. B*)
13. Native point defects and degradation of ZnSe-based devices; M. A. Berding, M. van Schilfgaarde, and A. Sher, (In preparation for submission to *Appl. Phys. Lett.*)
14. Bound states for anisotropic potentials and masses; R. S. Pfeiffer, Y.-J. Huang, and A.-B. Chen, *Phys. Rev. B* **48**, 8541 (1993).

APPENDIX A

Defect modeling studies in HgCdTe and CdTe

M. A. Berding, A. Sher, and M. van Schilfgaarde

J. Electron. Mater. **24**, 1129 (1995)

Defect Modeling Studies in HgCdTe and CdTe

M.A. BERDING, A. SHER, and M. VAN SCHILFGAARDE

SRI International, Menlo Park, CA 94025

We have used a quasichemical formalism to calculate the native point defect densities in $x = 0.22$ $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ and CdTe. The linearized muffin-tin orbital method, based on the local density approximation and including gradient corrections, has been used to calculate the electronic contribution to the defect reaction free energies, and a valence force field model has been used to calculate the changes to the vibration free energy when a defect is created. We find the double acceptor mercury vacancy is the dominant defect, in agreement with previous interpretations of experiments. The tellurium antisite, which is a donor, is also found to be an important defect in this material. The mercury vacancy tellurium antisite pair is predicted to be well bound and is expected to be important for tellurium antisite diffusion. We consider the possibilities that the tellurium antisite is the residual donor and a Shockley-Read recombination center in HgCdTe and suggestions for further experimental work are made. We predict that the cadmium vacancy, a double acceptor, is the dominant defect for low cadmium pressures, while the cadmium interstitial, a double donor, dominates at high cadmium pressures.

Key words: CdTe, defect complexes, defects, HgCdTe

INTRODUCTION

The pseudobinary semiconductor alloy $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ with $x = 0.22$ is currently the material of choice for high-performance detectors in the long-wavelength infrared (LWIR) (8–14 μm). Unlike most other II-VI systems, both extrinsic p- and n-type doping can be achieved in HgCdTe, although in as-grown material the electrical characteristics are often determined by native point defect concentrations. Understanding the properties of point defects and manipulation of their densities during growth and processing is essential to high-yield manufacturing of focal plane arrays (FPAs). As in other semiconductors, it is difficult to establish the presence and identity of all the important neutral and compensating point defects during growth and processing, much less to determine their concentrations. CdTe is important both as a substrate and passivating material for epitaxial lay-

ers of LWIR HgCdTe. Native point defects are of interest in CdTe in that they relate to its stoichiometry (which in turn has been shown to impact the minority carrier lifetimes¹ in HgCdTe devices), the formation and annihilation of tellurium precipitates,² and the self- and inter-diffusion coefficients that impact materials stability during growth, during subsequent processing, and over the device lifetime.

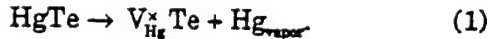
Our goal in this paper is to theoretically identify the important native defects in HgCdTe and CdTe, to calculate their densities as a function of growth and processing conditions, to validate our predictions by comparison with experimentally deduced properties of the native defects, and to suggest new experiments to begin to unravel the remaining mysteries in these materials. We have included in our analysis of HgCdTe both neutral and ionized states of eight native point defects and one defect pair. Our focus is on $x = 0.2$ (for comparison to annealing data),³ $x = 0.22$ (for LWIR applications), and $x = 1$ (for substrate and passivating layers) $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$. As we will show, we have at-

(Received October 4, 1994; revised January 20, 1995)

tempted to incorporate all of the important contributions to the defect formation free energies and adopt a first-principles approach for most of the quantities we calculate.

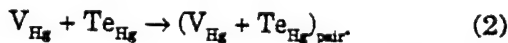
METHODOLOGY

To calculate the concentration of the native point defects in HgCdTe, we employ a quasichemical formalism⁴ in which the formation of independent crystalline defects is expressed as chemical reactions. For example, for the formation of the neutral mercury vacancy in HgTe with the mercury vapor as the reference phase, we have the reaction



Our notation is that of Kröger⁴ in which the primary symbol is the defect species and the subscript indicates the site the defect occupies; V indicates the vacancy, and I the interstitial, and no subscript indicates the species is occupying its normal lattice position; \times indicates the neutral defect species. Although we have chosen the mercury in the vapor phase as our reference and thereby will choose a mercury pressure to determine the chemical potential of mercury in the system, one could have chosen some other reference state for the mercury or tellurium (for example, by specifying a binary solution of $\text{Hg}_{1-y}\text{Te}_y$ at some temperature T in coexistence with the HgTe solid, as is essentially done during liquid phase epitaxy [LPE] growth). For our equilibrium calculations, the limits of mercury pressures within the existence region of $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ are taken from experiment.⁵

In a similar manner, reactions can be written for the other point defects of the system. In our analysis, we have included eight native point defects (plus their ionized species and the electron and hole): the mercury and tellurium vacancies (V_{Hg} and V_{Te}), the mercury and tellurium antisites (Hg_{Te} and Te_{Hg}), and two types of mercury and tellurium tetrahedral interstitials—one surrounded by four cation near-neighbors (Hg_{ttn} and Te_{ttn}) and the other surrounded by four anion near-neighbors (Hg_{itn} and Te_{itn}). We have also included the bound mercury vacancy tellurium antisite pair ($V_{\text{Hg}}\text{-Te}_{\text{Hg}}$)_{pair}, in which the vacancy and antisite occupy near-neighbor cations sites, via the reaction



For low densities of noninteracting defects, the law of mass action can be used to determine the neutral defect concentrations. For the mercury vacancy, this corresponds to

$$K_{V_{\text{Hg}}} \equiv \theta \exp\left(-\frac{F_{V_{\text{Hg}}}}{k_B T}\right) = [V_{\text{Hg}}^{\times}] \quad (3)$$

where $F_{V_{\text{Hg}}}$ is the reaction free energy for the neutral mercury vacancy, k_B is Boltzmann's constant, T is the temperature in kelvin, and θ is the number of unit cells per volume. Thus, once the reaction free energy

is known, it is straightforward to calculate the defect concentrations. Of course, all the work is involved in the calculation of the reaction free energies.

We have attempted to calculate all of the important contributions to the defect reaction free energies. An electronic contribution to the free energy results from the change in the total electron energy of the solid when a neutral defect is created; included in this energy is the electronic energy of free mercury atoms generated or consumed in the defect reaction. To calculate the electronic contribution, we employ the self-consistent first-principles full-potential linearized muffin-tin orbital method⁶ and the local-density approximation (LDA), including gradient corrections of the Langreth-Mehl-Hu type⁷ (which greatly improves the overbinding found in the LDA).^{8,9} The vibrational modes of the system are also changed when a defect is created; we calculate this change to the formation free energy (both the enthalpy and entropy) using a Green's function formalism within a valence force model plus point-charge ionic model. An entropy contribution to the formation free energy also arises from the partial occupation of degenerate levels associated with the defect and from the introduction of a preferred direction via a symmetry-lowering Jahn-Teller distortion.¹⁰ The combination of the electronic, vibrational, degeneracy, and translational (for the calculation of the chemical potential of the mercury atom in the vapor phase) free energies encompasses the primary contributions to the total defect formation free energies when referenced to a mercury vapor. Details of the results for these energies are given in Ref. 9. We have completed only preliminary calculations for the binding energy of the mercury vacancy tellurium antisite pair indicated in Eq. (3), using a 32-atom supercell and with overall lattice constant relaxation only. We expect that this defect may show further relaxation, which could lower the defect pair binding energy and therefore increase their density.

Although we are interested in studying the properties of point defects in HgCdTe, the neutral defect reaction free energies are calculated for HgTe; because we predict that the defects with highest concentrations (the mercury vacancy and the tellurium antisite) reside on the cation sublattice and therefore have four tellurium atoms for first nearest neighbors, we expect that the presence of cadmium will introduce a minor modification to the formation free energies, and of the order of other approximations made in the calculation (for example, using supercells). The presence of cadmium may have a larger impact on interstitial formation free energies, which can occupy sites with cations as near-neighbors. Although we find the density of interstitials to be relatively low,⁹ the cation interstitial in particular is very important in annealing of HgCdTe.¹² For the present, we have included the effects of the cadmium only by adjusting the number of sites available for mercury vacancies and in calculating of the band structure used in predicting the ionized point defect concentrations.

For further discussion and details of the calculation of the reaction free energies, see Ref. 9.

Because the native point defects will in general have localized levels in the band gap, we need to calculate the concentration of the ionized defects in addition to the neutral defects discussed above. Once the density of the neutral defects is determined, the concentration of ionized defects can be calculated from

$$\frac{[X']}{[X^*]} = \frac{g_{X'}}{g_{X^*}} \exp\left(\frac{\mu_F - E_a}{k_B T}\right) \quad (4)$$

for an acceptor defect X and

$$\frac{[Y^*]}{[Y^*]} = \frac{g_{Y^*}}{g_{Y^*}} \exp\left(\frac{E_d - \mu_F}{k_B T}\right) \quad (5)$$

for a donor defect Y. Once again, we have adopted the notation of Kröger: the bullet superscript indicates a positive charge and a prime, a negative charge. E_a and E_d are the one-electron acceptor and donor levels; μ_F is the Fermi energy, which is determined by demanding charge neutrality; and g_x is the degeneracy of the state, including both spin and Jahn-Teller¹⁰ splitting. Generalizations of these expressions are used for multiple ionized defects. The calculation of the electron and hole populations demands a knowledge of the temperature-dependent band structure and use of Fermi-Dirac statistics. The only significant empirical data we employ in this calculation are those needed to obtain the temperature-dependent band structure, which is extrapolated to high temperatures at which equilibration and annealing take place and for which we are going to predict defect concentrations. Further details are in Ref. 9.

DEFECTS IN HgCdTe

Equilibrium Native Point Defect Concentrations: Annealing Studies

Gibbs' phase rule tells us that for a system of three components (in our case mercury, cadmium, and tellurium) and two phases (zinc-blende solid and vapor), there are three degrees of freedom. In evaluating the equilibrium defect concentrations in $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$, we have chosen the temperature, the mercury pressure P_{Hg} , and the alloy composition x as these specified variables; the tellurium and cadmium pressures, the crystal stoichiometry, and the density of the various native point defects are determined by these parameters.

In Fig. 1, we show our results for the 77K hole concentrations in $x = 0.2 \text{ Hg}_{1-x}\text{Cd}_x\text{Te}$ as a function of mercury partial pressure for various high-temperature annealing conditions and compare them to the results of Vydyanath.^{3,13} We have assumed that the high-temperature defect structure is frozen in during a quench to 77K. All eight point defects discussed above have been included in our analysis. The agreement with experiment is quite good, given that the

only empirical data used were of the temperature dependence of the band structure used to calculate the intrinsic reaction constant. To demonstrate the sensitivity of our results to the free energies we are calculating from first principles, we have also shown in Fig. 1 the results of a calculation using a mercury vacancy formation energy that has been increased roughly 10%, plus a constant multiplicative constant of the low-temperature hole concentrations at all pressures and temperatures; as one can see, such minor modifications to our calculated parameters yield low-temperature hole concentrations that are in very good agreement with experiment. Certainly, refinements to our theory (for example, including a more accurate high-temperature band structure, more precise incorporation of alloy effects on formation free energies, anharmonic effects in the vibrational free energies, and going beyond the local density approximation) could account for discrepancies of this magnitude. In addition, the experimental data may be impacted by re-equilibration during the quench from high temperature.

At all temperatures, we predicted that equilibrated material will be p-type and that the dominant defect is the doubly ionized mercury vacancy, in agreement with mobility data.³ Our predictions differ from the analysis in Ref. 3 in several ways, though. First, although the data indicate that the material is intrinsic at all annealing temperatures, we predict that the material will be extrinsic for the higher annealing temperatures with $[h\cdot] \sim P_{\text{Hg}}^{-1/3}$, while at the lower annealing temperatures, the material will be intrinsic.

Figure 1
Box FPO

Fig. 1. Hole concentrations at 77K as a function of mercury partial pressure for material annealed at high temperatures, as indicated. Experimental data were taken from Vydyanath.³ Results of our theory are shown as solid lines. To demonstrate the sensitivity of our predictions to small changes in the calculated formation free energies, we show the dashed lines, which are the theoretical results, but with the mercury vacancy formation energy increased by 10% and including a rigid upward shift of the hole concentrations by a factor of 5.5. Note that the results shown here differ from those in Ref. 9 because the inclusion of the additional degeneracy factor for the singly ionized tellurium antisite.¹⁰

sic with $[h\bullet] \sim P_{\text{Hg}}^{-1}$ (for details see Ref. 9). This may be due to either a vacancy formation energy that is too small in our calculation (as demonstrated by the dashed line in Fig. 1) or errors in the high-temperature band structure, which we have extrapolated from the low-temperature formulas.^{14,15} Second, we find at higher temperatures and lower pressures, that the hole concentrations increase with increasing pressure, contrary to what would be expected due to mercury vacancy acceptors alone. In our theoretical predictions, this is due to compensation from tellurium antisites, which we predict are donors, and which are the second most dominant native point

defect in $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$. Because the equilibrium antisite density varies roughly as P_{Hg}^{-2} , compared to P_{Hg}^{-1} to $P_{\text{Hg}}^{-1/3}$ for the mercury vacancy, it is most important at low mercury pressures. At the highest temperatures, there does appear to be a corresponding role-off in the experimental data at the lower pressures, which may be indicative of compensation by tellurium antisites, but it could also be an experimental artifact caused by quenching inefficiencies for the higher temperatures.

Low-temperature mercury-saturated anneals are of technological importance to reducing the as-grown p-type carrier concentrations or to convert the material to n-type in nominally undoped material. In Fig. 2, we have plotted the concentrations of mercury vacancies and tellurium antisites as a function of inverse temperature for constant pressures within the existence region; neither of these concentrations include the defects that are bound into mercury vacancy tellurium antisite pairs. The lower boundaries correspond to the defect densities for mercury-saturated anneals, the upper boundary to tellurium-saturated anneals. As one can see, at a given pressure and temperature, the mercury vacancy concentrations are always in excess of the tellurium antisite population, and in equilibrium the material should always be p-type. As discussed in the following section, we expect that full equilibration of the tellurium antisite density may not take place at lower annealing temperatures, so that the equilibrium concentrations predicted for this defect may not always be realized.

Annealing schedules are also of technological importance in forming p-n junctions in as-grown p-type material via mercury in diffusion.^{16,17} The results of these experiments depend on the mechanism by which mercury diffuses and are related to the identity of the residual donor, both of which will be discussed further below.

The Tellurium Antisite in $\text{Hg}_{0.75}\text{Cd}_{0.25}\text{Te}$

Having predicted that the tellurium antisite will be

Figure 2
Box FPO

Fig. 2. Total mercury vacancy (solid line) and tellurium antisite (dashed line) densities as a function of mercury partial pressures (atm). Full equilibration of all defects is assumed at all temperatures (see text for further discussion).

Figure 3a
Box FPO

Figure 3b
Box FPO

Figure 3c
Box FPO

Fig. 3. The concentration of native points as a function of mercury pressure within the phase stability region at (a) 500°C, corresponding to the LPE growth temperature, (b) 185°C, the growth temperature for MBE, and (c) 220°C, a typical temperature for mercury-saturated anneals. A range of concentrations for the mercury vacancy tellurium antisite pair is shown, based on our preliminary results, as discussed in the text.

present in substantial densities, we now address the evidence for their existence in LWIR HgCdTe. We have predicted that the tellurium antisite is a donor, although we have not yet extended our theory to predict whether it is a single or double donor, nor the precise location of the defect levels in the gap. This is perhaps the most difficult part of our calculation because of the LDA band gap errors intrinsic to our method and the dispersion arising from the use of supercells that give defect-level widths larger than the $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ band gap. To establish the presence of tellurium antisites without a prediction of the donor level(s) in the gap, we must consider the possibility that there are resonant level(s), shallow level(s), deep level(s), or some combination of these, associated with the tellurium antisite, and look for evidence supporting the presence of the antisite.

One of the outstanding questions in the properties of HgCdTe is the identification and elimination of the residual donor that controls the carrier concentration in material annealed at low temperatures under mercury-saturated conditions. Although purification of starting materials has led to a lowering of the residual donor density, a lower limit of $\sim 10^{14} \text{ cm}^{-3}$ seems to have been reached. To examine the possibility that the tellurium antisite is the residual donor, we have calculated the defect densities at 500°C , roughly the LPE growth temperature from both the mercury- and tellurium-rich melts, at 185°C , the typical molecular beam epitaxy (MBE) growth temperature, and at 220°C , roughly where low-temperature mercury-rich anneals are done; results are shown in Fig. 3a–3c. It is clear that the material is expected to be p-type for both MBE- and LPE- grown material, and that although postgrowth, mercury-saturated anneals (the right side of Fig. 3c) will lower the vacancy density below the 10^{13} cm^{-3} level, the material is still predicted to be mercury-vacancy-doped p-type. We thus conclude that the residual donor is not due to an equilibrium concentration of native point defects.

The possibility remains that the residual donor is associated with a nonequilibrium defect concentration of tellurium antisites. We expect that the diffusion coefficient of the tellurium antisite is relatively small because the diffusion of an antisite will necessarily involve at least one additional point defect, such as the mercury vacancy or the tellurium interstitial (diffusion via a consorted exchange of a tellurium antisite with a mercury atom on an adjacent cation lattice site is unlikely). As such, the tellurium antisite may not reach equilibrium densities for the times and temperatures corresponding to the low-temperature mercury-saturated anneals typically employed to equilibrate the mercury vacancy density (Fig. 3c). If tellurium antisite densities are in fact equilibrated at a temperature at which the antisite diffusion effectively stops during cooldown from the growth temperature, then the antisites may be frozen in at higher, nonequilibrium concentrations corresponding to the residual donor density.

Addressing the question as to why the antisite

density might be frozen in at roughly the same concentration for LPE material grown from both the mercury and tellurium melt, we return to the means by which tellurium antisite diffuses in the material, and assume it diffuses by a mercury vacancy mechanism. This assumption is motivated by several factors. First, this mechanism involves only one point defect in addition to the tellurium antisite and involves a simple migration mechanism between the two defects. Second, it involves the mercury vacancy, whose density is fairly high, and therefore the probability of tellurium antisite mercury vacancy pairs is expected to be fairly high. In addition, the mercury vacancy is an acceptor and the tellurium antisite is a donor so they are expected to form a bound pair based on Coulombic attraction, and the mercury vacancy is too small for the lattice, while the tellurium antisite is too large, so that there is a mechanical attraction between them as well. Finally, the migration mechanism involving an interstitial—for example, via a kickout mechanism whereby a mercury interstitial kicks out the tellurium antisite to form a tellurium interstitial—is unlikely since it involves defects that are all too large for the lattice and are donors and therefore are unlikely to form pairs. The kickout mechanism may be important when mercury interstitials are injected into the material, and is discussed later in this paper. The likelihood of the tellurium antisite diffusion proceeding by the mercury vacancy mechanism is further supported by our preliminary prediction of the mercury vacancy tellurium antisite binding energy of 1.1 eV; the corresponding densities are shown in Fig. 3. We show our predictions for the pair density as a range in which the lower limit corresponds to the defect pair being neutral, as our preliminary predictions indicate, and the upper limit corresponding to the pair having a donor state at the valence band edge and an acceptor level at the conduction band edge. Unlike the native point defects, we expect that these pair defect concentrations may change as we refine the free energy calculations. If tellurium antisite diffusion is via the mercury vacancy, the diffusion coefficient will be proportional to the density of the mercury vacancy tellurium antisite pairs, that is

$$D_{\text{Te}_{\text{Hg}}} \sim \left[(V_{\text{Hg}} + \text{Te}_{\text{Hg}})_{\text{pair}} \right]. \quad (6)$$

From Fig. 3a for LPE growth, one can see that the density of defect pairs for low mercury pressures, corresponding to material grown from the tellurium-rich melt, is four orders of magnitude higher than for that grown from the mercury-rich melt. This implies that although material grown from the tellurium melt will contain higher densities of tellurium antisites, in the cooldown from the growth temperature the tellurium antisites will re-equilibrate much more rapidly due to the large pair density than will material grown from the mercury-rich melt. In contrast, for material grown from the mercury melt, the tellurium antisites present at the growth tempera-

ture, although lower, may be frozen in because the pair density is also lower.

We now turn to growth by MBE, which takes place at 185°C under mercury-poor conditions, corresponding to the left side of Fig. 3b. At the phase stability boundary, we predict that the material will be mercury-vacancy doped p-type, with a carrier concentration of $\sim 10^{15} \text{ cm}^{-3}$. Experimentally as-grown material is found to be either n- or p-type with carrier concentrations in the 10^{15} cm^{-3} range.¹⁸ Because MBE growth is a nonequilibrium process, it is possible that growth may take place beyond the existence region; for example, if the equivalent pressure were of the order of 10^{-7} atm at 185°C, the materials would be tellurium antisite doped. While this provides a possible explanation of how the material could be n- or p-type as-grown based on equilibrium concentrations of defects, extending this argument, one would expect under some growth conditions to be able to obtain highly compensated material with very low carrier concentrations; this is never seen, to our knowledge. A more likely possibility is that nonequilibrium densities of point defects are frozen into the MBE material resulting from details of the surface kinetics.

There is substantial experimental evidence of deep levels in LWIR HgCdTe, which have not yet been definitively associated with any particular defect. The minority carrier lifetimes in vacancy-doped material are limited by Shockley-Read recombination; deep-level transient spectroscopy (DLTS) measurements¹⁹⁻²¹ have identified two donor-like defect levels at $0.4E_{\text{gap}}$ and $0.75E_{\text{gap}}$. The densities of these levels roughly track the mercury vacancy concentration, although substantial scatter in the correlation is observed.¹⁹ Neither of these levels is believed to be associated with the mercury vacancy itself, nor do they appear to be associated with the same defect because they do not track one another. In equilibrium, the tellurium antisite density will track with the mercury vacancy concentration, although it will show a sharper dependence on the mercury pressure. In addition, as discussed above, we do not expect that equilibrium concentrations of tellurium antisites will be present except at very high temperatures corresponding to liquid-phase growth, and thus the ratio of tellurium antisites to mercury vacancies expected in equilibrium may not be experimentally realized, leading to substantial deviations from the equilibrium ratio and scattering in their concentrations. Thus, it is plausible that one of the Shockley-Read recombination centers is associated with the tellurium antisite, although to confirm this possibility a more quantitative prediction of the ionization levels of the antisite is needed. Deep levels associated with the tellurium antisite may also be responsible for the $1/f$ noise, which is found to be roughly proportional to the mercury vacancy concentration²² or they may enhance interband tunneling and thereby contribute to dark currents.

Diffusion

Self-diffusion of mercury is important for the annealing of as-grown mercury-vacancy doped p-type material to n-type, to form p-n junctions,¹⁶ and for understanding junction stability in HgCdTe devices. Although both the mercury vacancies and interstitials are mobile and contribute to mercury diffusion,²³ it is the interstitial diffusion that is found to dominate in the modeling of low-temperature anneals.^{12,17} Thus, although we find equilibrium mercury interstitial concentrations that are negligible in terms of their contribution to the net carrier densities,⁹ they will be important to mercury transport in the material, and therefore their properties are of interest. For both mercury vacancy and interstitial diffusion, more than just defect concentrations enter into the determination of the diffusion coefficients; for the present discussion, we will address only how our defect concentrations relate to the measured diffusion coefficients.

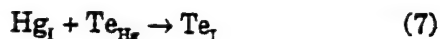
We predict that mercury interstitials are donors and in equilibrium are present in concentrations that are less than 10^6 cm^{-3} at 250°C.⁹ We have compared our results with those of annealing simulations by the Stanford group¹² and find that the concentration of interstitials we have predicted at $\sim 200^\circ\text{C}$ are several orders of magnitude too small to account for their modeling of the formation of p-n junctions. Although a number of approximations in our calculation of the formation energies will affect our interstitial formation free energy (for example, the use of supercells and approximations of the ionization energies), we do not expect these to account for this large a discrepancy. As discussed in the Methodology section above, we have completed the calculation of the electronic contribution to the formation free energies using pure HgTe. For the mercury vacancy and the tellurium antisite that are the major defects in HgCdTe and that occupy the cation sublattice and therefore are surrounded by four tellurium atoms, this is probably not such a bad assumption. Corrections to the electronic energy due to the presence of cadmium may be larger for the interstitials that see four cations as their first nearest neighbors in one tetrahedral site and six cations as near second neighbors in the other tetrahedral site. We are currently calculating the correction to our electronic energies, taking explicit account of cadmium in the lattice to see if it will eliminate the discrepancy with the Stanford model predictions.¹²

There has been a recent investigation²⁴ in $x = 0.22$ and $0.24 \text{ Hg}_{1-x}\text{Cd}_x\text{Te}$ on samples in which high concentrations of nonequilibrium mercury interstitials have been introduced. The observed deep levels near 45 and 60 meV above the valence-band edge were argued to be associated with the mercury interstitials. As discussed above, at this point our calculations are not able to determine the precise positions of the defect levels in the gap, but rather have predicted only that the mercury interstitials will be donor-like. If in fact there are donor levels associated with the mercury

The Role of Native Point Defects in Self-

interstitials that are near midgap, we would expect them to exhibit a series of levels corresponding to the different alloy environments about an interstitial. In addition, there are two classes of interstitials that we expect to have relatively high equilibrium densities, both of which occupy tetrahedral sites; the first is surrounded by four cation nearest neighbors, and the second is surrounded by four anion nearest neighbors. Although it is tempting to associate the two defect levels experimentally observed with these two classes of interstitial sites, there is no apparent reason why two levels are seen in the $x = 0.22$ material and only one level in for $x = 0.24$.

Finally, we return to the discussion of the mechanism for tellurium antisite diffusion and re-address the knockout mechanism that proceeds via the reaction



Our calculations predict that this is an exothermic reaction, with an energy of ~ 0.8 eV (the entropy gained in having two point defects rather than just one must also be considered in establishing the equilibrium concentrations of these defects). In nonequilibrium situations in which mercury interstitials are introduced into the material—for example, during ion-beam milling or oxide baking²⁵—this reaction will be pushed to the right, and excess tellurium interstitial will be produced. Thus, in such situations, one might expect tellurium antisite diffusion via the knockout mechanism to be a stronger competitor to the mercury vacancy mechanism, although the barriers to the formation of $\text{Hg}_i\text{-Te}_{\text{Hg}}$ pairs still exist, as discussed above.

In presenting the tellurium antisite as a candidate for the residual donor, we argued why its density might be fixed in the n-type material. On the other hand, when discussing the possibility that it is related to a Shockley-Read recombination center in vacancy-doped p-type material, we argued why its density might vary, depending on the cooldown rate and so forth. These two arguments are somewhat inconsistent. Although the possibility still exists that the tellurium antisite is both the residual donor (via a first ionization level that resonates in the gap) and a Shockley-Read center (via a midgap second ionization level), to be convincing a firmer correlation between the two would have to be established.

The Mercury Vacancy Tellurium Antisite Pair

Several additional consequences of the presence of the mercury vacancy tellurium antisite pairs should be discussed. As one can see from Fig. 3, we are predicting a very large concentration of the defect pairs, which may even exceed the mercury vacancy concentrations for LPE material grown from the tellurium-rich melts. If the pair is electrically inactive, as our preliminary calculations predict, it will not impact the carrier concentrations or mobility. Such a large density of pairs does imply that the nonstoichiometry of the material is larger than that due to the vacancy concentrations, particularly at

lower temperatures where we predict that the equilibrium concentrations of neutral pairs will approach that of the mercury vacancy.

Note that the large binding energy of the mercury vacancy tellurium antisite pair suggests that other bound pairs may be present in the material. For example, a bound Frenkel defect (involving the mercury vacancy-mercury interstitial pair) that involves an acceptor and a donor defect with opposite lattice strains may be important and will impact diffusion of mercury in the lattice. The most likely consequences will be to increase the annihilation capture cross section of mercury interstitials into mercury vacancies and to present a barrier to the formation of free Frenkel defects through a geminate process. These phenomena will be important to understanding diffusion in HgCdTe.¹² Evidence for the mercury vacancy substitutional indium pair has been seen using a nuclear hyperfine technique.²⁶ This defect is similar to the mercury vacancy tellurium antisite pair, and thus we expect it may be well bound. Both of these pairs merit further investigation.

DEFECTS IN CdTe

Like HgCdTe, CdTe has a wide stability region and can be doped both p- and n-type. Its use as both a substrate (along with $\text{Cd}_{1-x}\text{Zn}_x\text{Te}$)²⁷ and a passivant make it an important material in the manufacturing of HgCdTe LWIR FPAs.

We have predicted the density of neutral native point defects in CdTe as a function of temperature and pressure; results for 700°C are shown in Fig. 4. We predict the same electrical type (donor vs acceptor) for the native point defects as in HgCdTe; for example, we find that the cadmium vacancy is an acceptor and the cadmium interstitial and tellurium antisite are

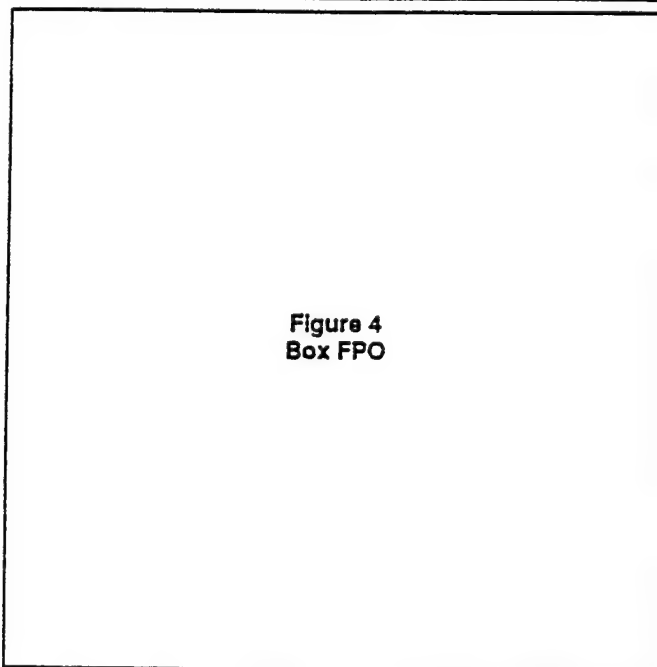


Figure 4
Box FPO

Fig. 4. Neutral native point defect densities of CdTe within the stability region at 700°C.

donors. Based on the neutral native point defects, we expect that CdTe solid can exist with both excess cadmium and excess tellurium. We find that the cadmium interstitial is the dominant defect for high cadmium pressures, and the cadmium vacancy is most important at mid and low cadmium pressures.

Because the native point defects may have energy levels in the gap associated with them, it is important to include these in the analysis. We have begun to calculate the ionization levels associated with the native point defects but have not yet included the Jahn-Teller distortions,²⁸ which can significantly alter the localized energy levels. For this discussion, we shall assume that the cadmium vacancy has a first ionization level near the valence-band edge and a second level near midgap²⁹ and assume that the cadmium interstitial is a double shallow donor.² In this case, we predict a p- to n-type conversion at 700°C at relatively high cadmium pressures, in agreement with experiment.²

From Fig. 4, we see that the tellurium antisite becomes more important as the cadmium pressure is lowered. Depending on the energies associated with the tellurium antisite donor levels, the material may be highly compensated at the lowest pressures, or another p- to n-type conversion may even occur. Experiments designed to equilibrate on the cadmium-poor side of the stability region can be used to test for the presence of tellurium antisites through both their impact on electrical activities and the presence of localized levels.

CONCLUSIONS AND SUGGESTED EXPERIMENTS

The theory developed here clearly indicates that in $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$ equilibrated at typical annealing temperatures and pressures (the right side of Fig. 3c), the tellurium antisite and mercury vacancy tellurium antisite pair densities are well below levels that can impact device performance. The primary outstanding question is whether or not tellurium antisite diffusion rates are high enough that normally processed samples fully equilibrate. Experiments are needed to modify these defect populations in a controlled manner so their impact on carrier concentrations, lifetimes, and other device properties can be determined. The basic idea is to modify the tellurium antisite concentration by choosing annealing temperatures, mercury pressures, and times that are long enough to permit a measurable portion of samples to equilibrate.

First, we consider annealing experiments to test for the possibility that the antisite is the residual donor. If tellurium antisite diffusion is so slow that it is not equilibrated during low-temperature mercury-saturated anneals, one must first anneal at higher temperatures and lower mercury pressures to introduce tellurium antisite mercury vacancy pairs (this step could be eliminated for LPE material grown from the tellurium melt), followed by a lower-temperature, low-mercury-pressure anneal to reduce the antisite density, while still maintaining a relatively high

density of pairs, concluded by a low-temperature, mercury-saturated anneal to reduce the mercury vacancy concentration even further. A series of experiments in which the conditions of the first two anneals were varied could be designed to test for the effect they have on the residual donor concentration. A similar experiment could be performed to correlate the Shockley-Read center in mercury-vacancy-doped material with nonequilibrium tellurium antisite densities.

Although we have proposed a series of experiments to establish whether the tellurium antisite is the residual donor or a Shockley-Read recombination center based on nonequilibrium densities, one may be able to design a series of experiments in which equilibrium populations of tellurium antisites are obtained by choosing high enough temperatures, thin enough samples, and long enough annealing times. One could choose annealing conditions to manipulate the mercury vacancy and tellurium antisite populations independently. For example, a 300°C anneal to reduce the hole concentration to $\sim 10^{18}$ will result in an order-of-magnitude more tellurium antisites than an anneal at 400°C to achieve the same hole concentration; Fig. 2 can be used in guiding such a study. The correlation of the tellurium antisite densities with the Shockley-Read center should be possible by doing minority carrier lifetime and/or DLTS measurements on materials with the same mercury vacancy hole concentrations achieved by anneals at different temperatures. Although a study of this type was recently presented,¹ and for a given hole concentration a correlation of lifetimes with the annealing temperature was observed, the effective annealing conditions there were set by the stoichiometry of CdTe cap layers and thus were more complicated than a simple anneal with a controlled mercury overpressure.

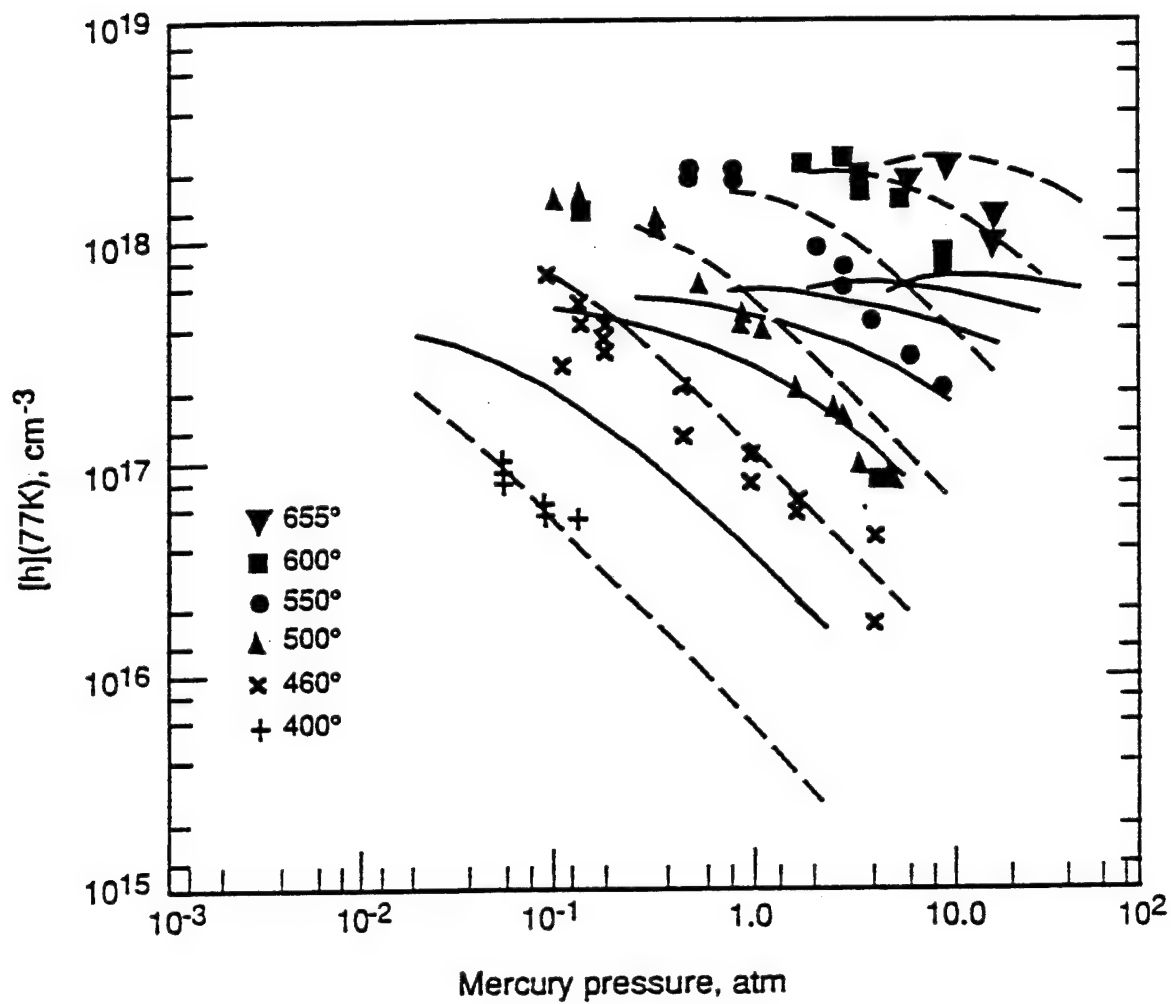
Experiments to test the presence of the mercury vacancy tellurium antisite pair are similar to those proposed above for the isolated tellurium antisite. Our preliminary calculations indicate that the pair will be electrically inactive and have no states in the band gap. As such, we do not expect their presence to have a direct impact on the electrical properties, but their presence will be manifested in their impact on tellurium antisite diffusion. Note that these experiments proposed to test for the presence of the tellurium antisite cannot discriminate between the isolated tellurium antisite and the mercury vacancy tellurium antisite pair.

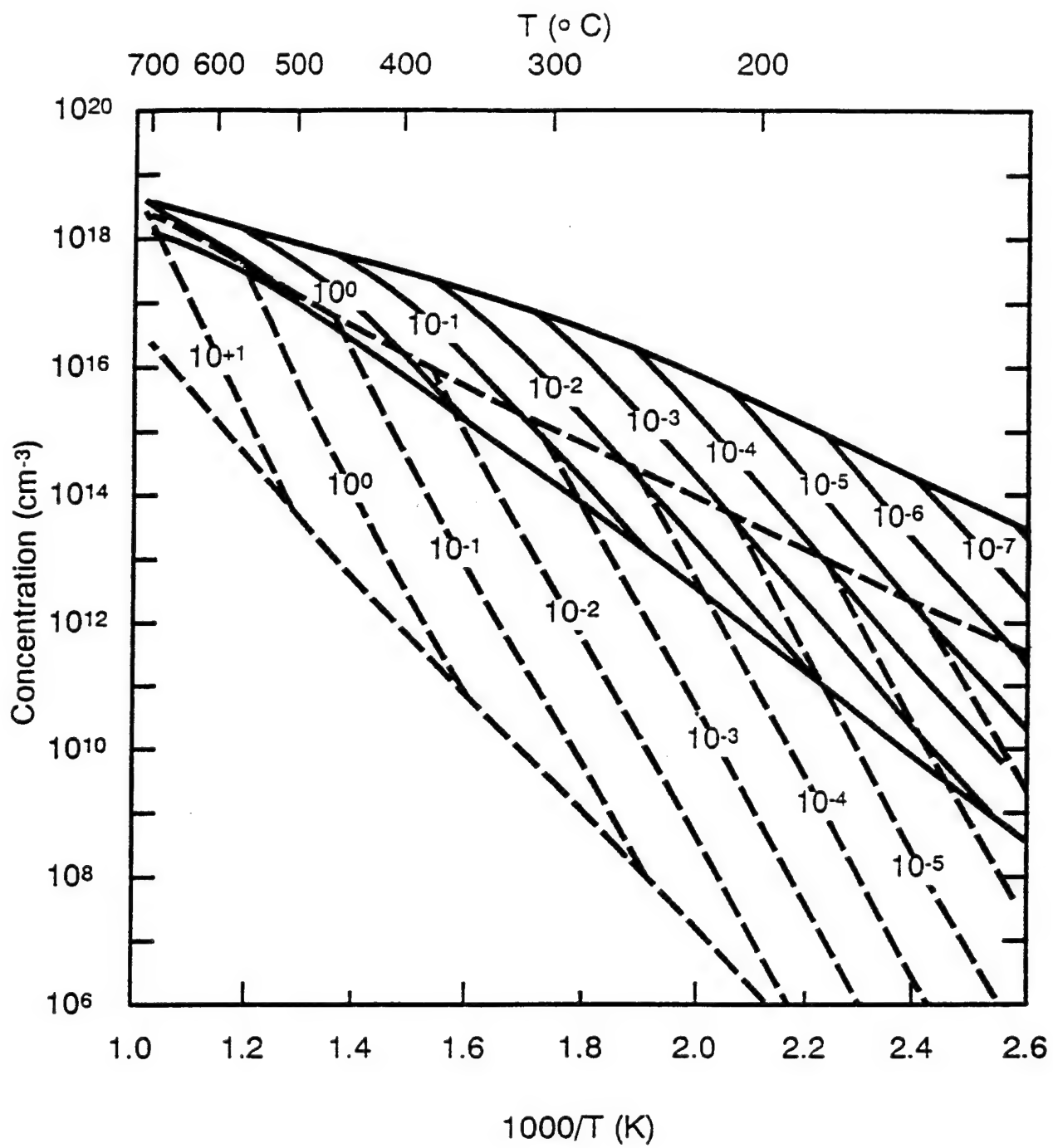
ACKNOWLEDGMENT

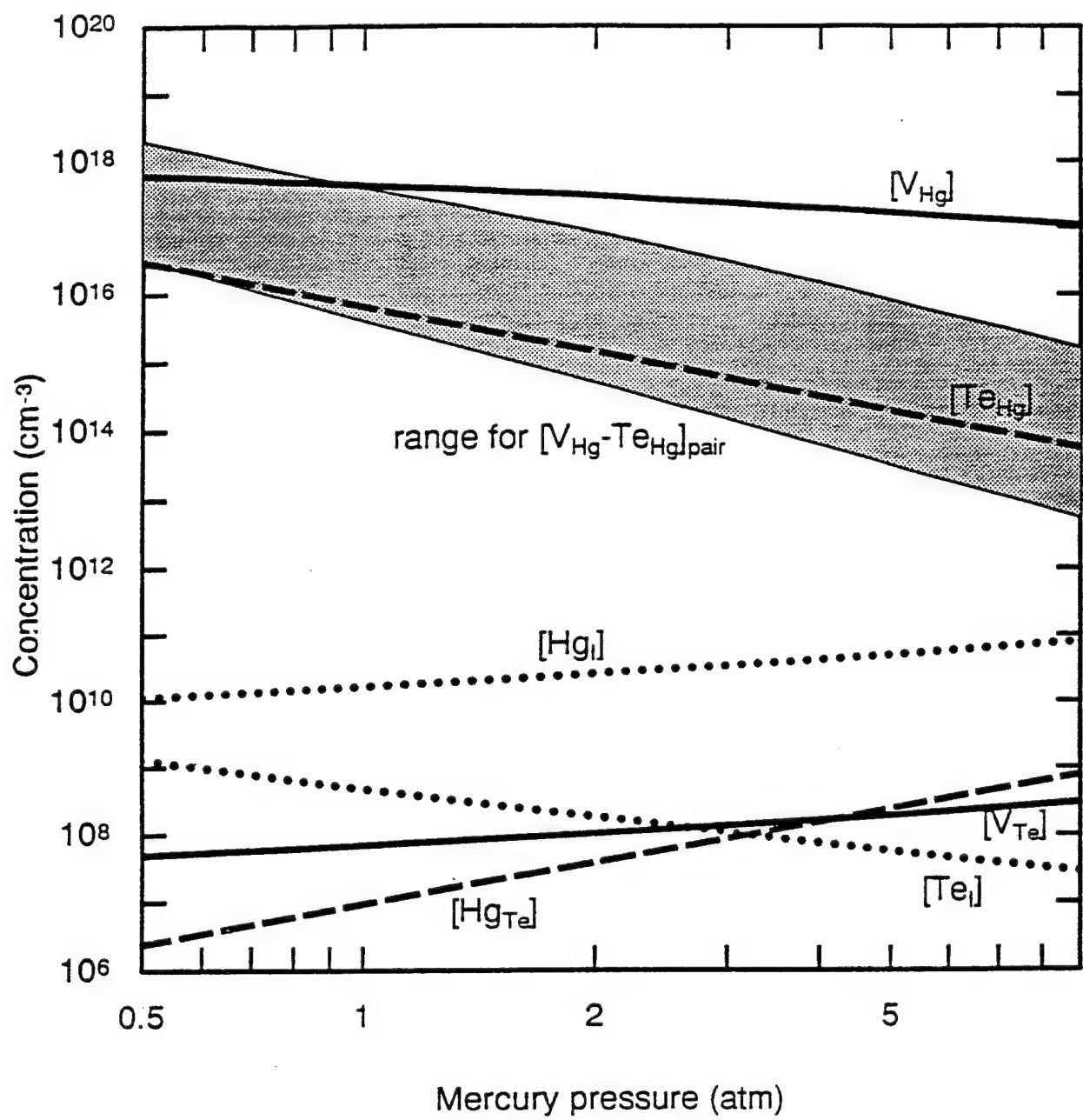
The work reported here was supported by NASA Contract No. NAS1-18226, ONR Contract No. N00014-89-K-132, and ARPA Contract No. MDA972-92-C-0053. Use of the Numerical Aerodynamic Simulation supercomputer facilities at NASA Ames Research Center is gratefully acknowledged.

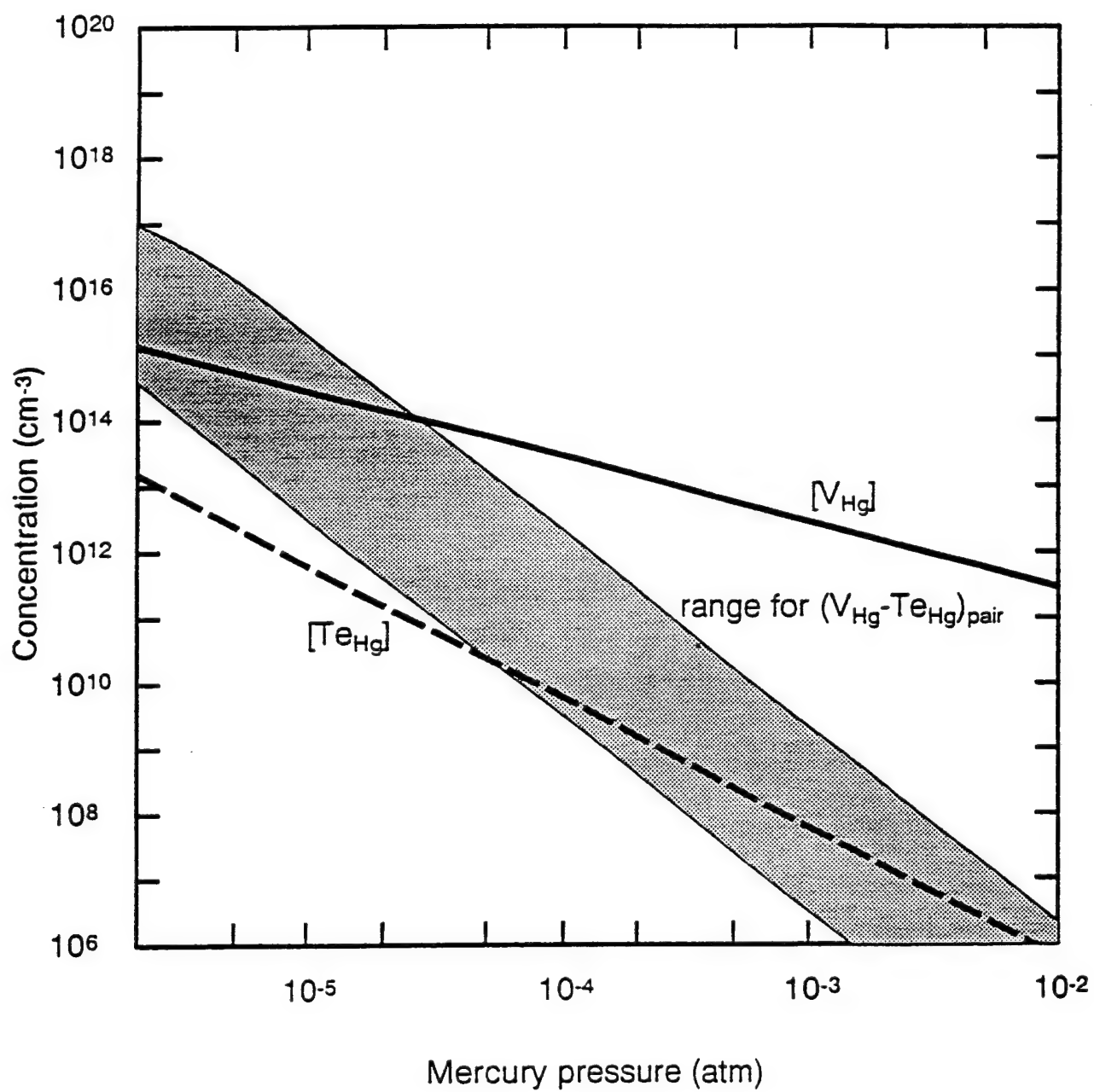
REFERENCES

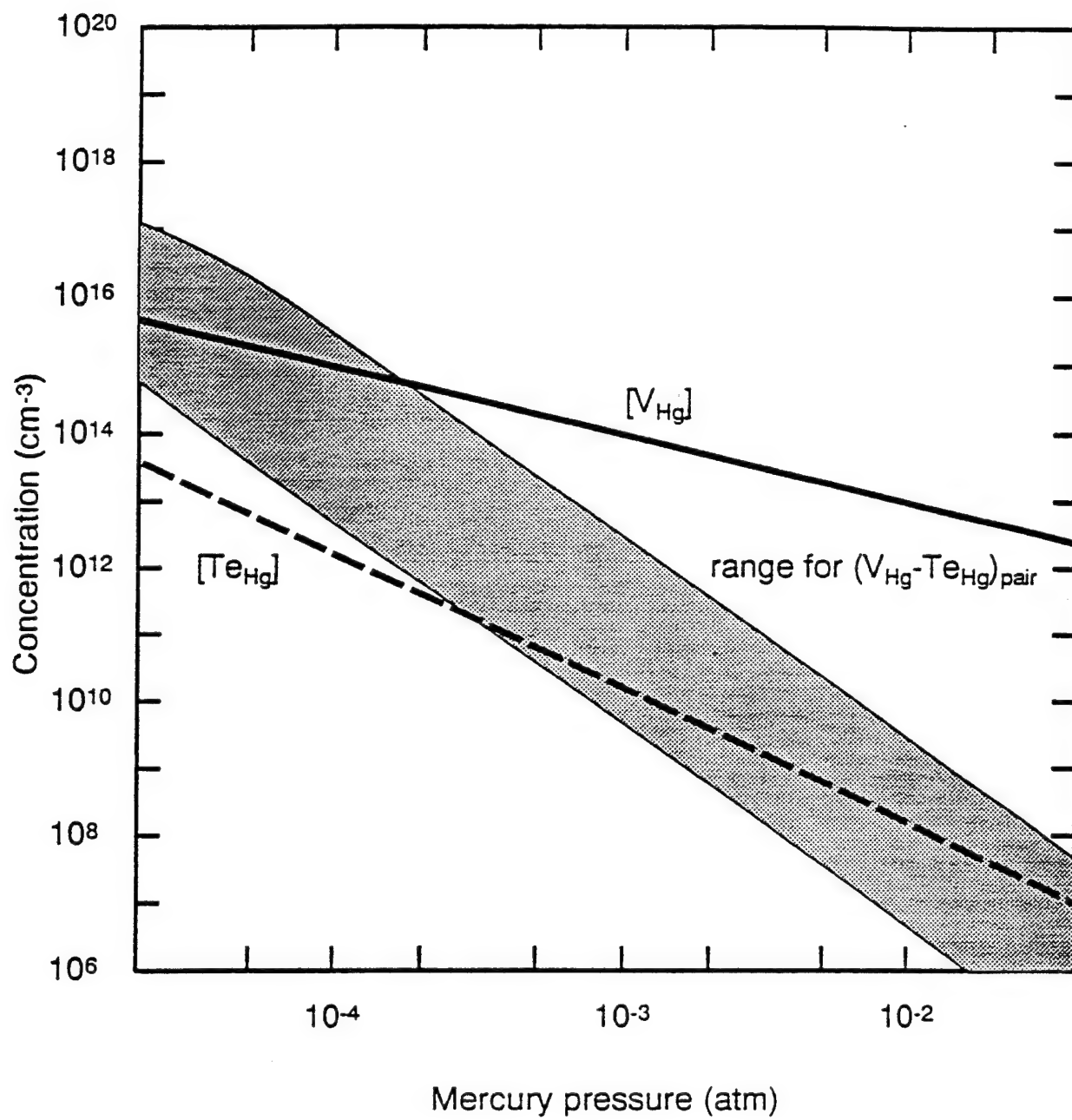
1. C.F. Wan, M.C. Chen, J.H. Tregilgas, T.W. Orent and J.D. Luttmer, 1993 MCT Workshop.
2. H.R. Vydyanath, J. Ellsworth, J.J. Kennedy, B. Dean, C.J. Johnson, G.T. Neugebauer, J. Sepich and P.-K. Liao, *J. Vac. Sci. Technol. B* 10, 1476 (1992).
3. H.R. Vydyanath, *J. Electrochem. Soc.* 128, 2609 (1981).
4. F.A. Kröger, *The Chemistry of Imperfect Crystals* (New York: J. Wiley & Sons, Inc., 1964).
5. T. Tung, M.H. Kalisher, A.P. Stevens and P.E. Herning, *Mat. Res. Soc. Symp. Proc.* 90, 321 (1987) and references therein.
6. O.K. Andersen, O. Jepsen and D. Glotzel, *Highlights of Condensed Matter Theory*, ed. F. Bassani et al. (Amsterdam, The Netherlands: North Holland, 1985), p. 59.
7. D. Langreth and D. Mehl, *Phys. Rev. B* 28, 1809 (1983).
8. M. van Schilfgaarde, A.T. Paxton, M.A. Berding and M. Methfessel (in preparation).
9. M.A. Berding, M. van Schilfgaarde and A. Sher, *Phys. Rev. B* 50, 1519 (1994).
10. In Ref. 9, we assumed degeneracies of one, two, and one for the neutral, singly, and doubly ionized states of all donor and acceptors in HgCdTe. This is valid if only the α_c states are occupied and the triply (sixfold including spin) degenerate t_2 states remain unoccupied. (Recall that the point group of an undistorted point defect is T_d .) If the t_2 states are partially occupied, an additional degeneracy will be present that should be included in our calculations; this degeneracy was not included in Ref. 9 but will be added in this paper. For an undistorted defect with the full T_d symmetry, this degeneracy arises from the partial occupation of the sixfold degenerate state (constrained to have a maximal number of spins paired); if the local symmetry is lowered due to a Jahn-Teller distortion (to point group D_{2d} for one or two electrons in the t_2 state, to point group C_{2v} for three, four, or five electrons in the t_2 state),¹¹ this additional degeneracy arises from the extra degree of freedom present in the introduction of a locally preferred direction in space. Going through the counting, we obtain degeneracy factors of (6,3,6,3,6,1) for one to six electrons in the t_2 state, respectively. Results reported here include these full degeneracy factors.
11. See, for example, the discussion of symmetry-lowering distortions in silicon in G.A. Baraff, E.O. Kane and M. Schlüter, *Phys. Rev. B* 21, 5662 (1980).
12. S.L. Holander and C.R. Helms, private communication.
13. H.R. Vydyanath and C.H. Hiner, *J. Appl. Phys.* 65, 3080 (1989).
14. G.L. Hansen and J.L. Schmit, *J. Appl. Phys.* 54, 1639 (1983).
15. G.L. Hansen, J.L. Schmit and T.N. Casselman, *J. Appl. Phys.* 53, 7099 (1982).
16. H.F. Schaake, J.H. Tregilgas, J.D. Beck, M.A. Kinch and B.E. Gnade, *J. Vac. Sci. Technol. A* 3, 143 (1985); H.F. Schaake, *J. Electron. Mater.* 14, 513 (1985).
17. J.L. Meléndez and C.R. Helms, *J. Electron. Mater.* 22, 999 (1993).
18. See, for example, R. Sporken, M.D. Lange, S. Sivanathan and J.P. Faurie, *Appl. Phys. Lett.* 59, 81 (1991).
19. C.E. Jones, V. Nair and D.L. Polla, *Appl. Phys. Lett.* 39, 248 (1981).
20. C.E. Jones, V. Nair, J. Lindquist and D.L. Polla, *J. Vac. Sci. Technol.* 21, 187 (1982).
21. C.E. Jones, K. James, J. Marz, R. Braunstein, M. Burd, M. Eetemadi, S. Hutton and J. Drumheller, *J. Vac. Sci. Technol. A* 3, 131 (1985).
22. R. Schiebel and D. Bartholomew, *J. Electron. Mater.* 24, XXX (1995).
23. See, for example, D.A. Stevenson and M.-F.S. Tang, *J. Vac. Sci. Technol. B* 9, 1615 (1991).
24. C.L. Littler, E. Maldonado, X.N. Song, Z. Yu, J.L. Elkind, D.G. Seiler and J.R. Lowney, *J. Vac. Sci. Technol. B* 10, 1466 (1992).
25. J.L. Elkind, *J. Vac. Sci. Technol. B* 10, 1460 (1992).
26. W.C. Hughes, M.L. Swanson and J.C. Austin, *Appl. Phys. Lett.* 59, 938 (1991); *J. Electron. Mater.* 22, 1011 (1993).
27. Because of the low zinc concentrations, our results should be applicable to $x = 0.04$ Cd_{1-x}Zn_xTe, which is the more recent substrate of choice for LWIR Hg_{1-x}Cd_xTe devices.
28. The Jahn-Teller distortion most certainly exists at low temperatures. At high temperatures corresponding to liquid phase growth, these distortions may get washed out by thermal vibrations in the lattice. We are currently examining this issue.
29. P. Höschi, R. Grill, J. Franc, P. Moravec and E. Belas, *Mat. Sci. and Engr.* B16, 215 (1993).

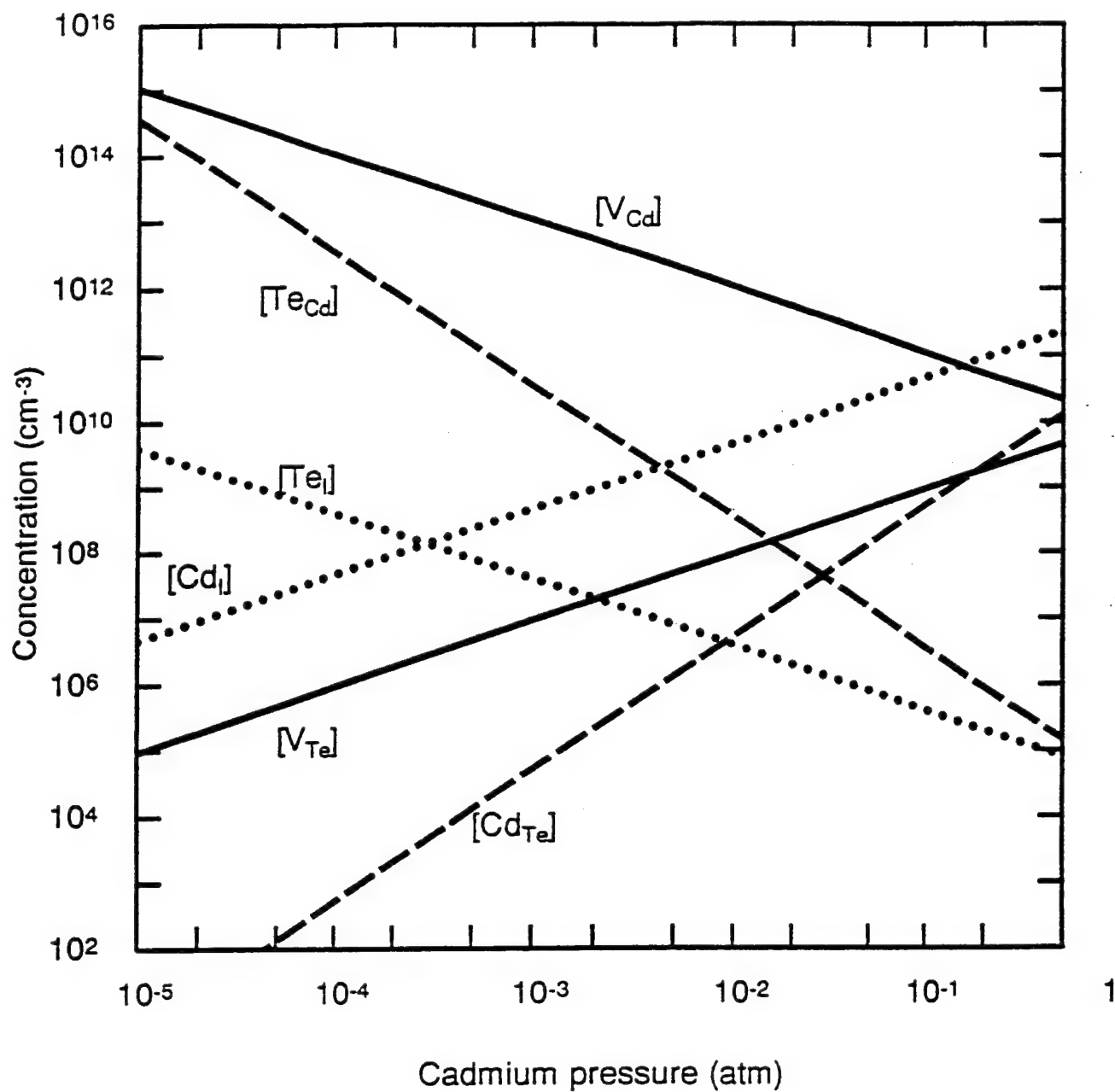












APPENDIX B

First-principles calculation of native defect densities in $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$

M. A. Berding, M. van Schilfgaarde, and A. Sher

Phys. Rev. B **50**, 1519 (1994)

First-principles calculation of native defect densities in $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$

M. A. Berding, M. van Schilfgaarde, and A. Sher

SRI International, Menlo Park, California 94025

(Received 3 March 1994)

We use a quasichemical formalism to make quantitative predictions of the native point defect densities in $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$. The electronic contribution to the defect-formation free energy is calculated using the self-consistent first-principles full-potential linearized muffin-tin orbital method and the local-density approximation (LDA). A gradient correction is added to the LDA result so that absolute reference to the chemical potential of the mercury vapor phase can be made. A Green's function approach based on a valence force field plus a point Coulomb model is used to calculate the vibrational contributions to the defect free energy (both energy and entropy). We find the double acceptor mercury vacancy is the dominant defect, in agreement with previous interpretations of experiments. The tellurium antisite is also found to be an important defect in this material. Predictions of the low-temperature hole concentrations are made as a function of annealing temperature and compared with available experiments. The order of magnitude of our predictions agrees well with experimental results, and discrepancies can be attributed to contributions to the free energy that we have neglected or to inaccuracies in the intrinsic reaction constant used. Suggestions for further experimental work are made.

I. INTRODUCTION

The pseudobinary semiconductor alloy $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ with $x=0.22$ is currently the material of choice for high-performance detectors in the long-wavelength infrared (8–14 μm). Unlike other II-VI systems, both extrinsic p - and n -type doping can be achieved in $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$, although in as-grown material the electrical properties are often determined by native point defect concentrations. The dominant defect is believed to be a double-acceptor mercury vacancy;¹ post-growth low-temperature mercury-saturated anneals are routine for the reduction of the mercury vacancy concentration. As in other semiconductors, it is more difficult to establish the presence and identity of neutral and compensating point defects, much less to determine their concentrations. Diffusion measurements² indicate the presence of mercury vacancies and mercury interstitials, as well as tellurium interstitials, although no unambiguous determination of their densities can be made from these experiments.

Although extended defects such as dislocations often appear to be the performance limiter in current state-of-the-art $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$ devices,^{3–5} a number of mysteries still persist that may relate to native point defects. (1) For operation at 40 K, there is a variation in R_0A_j and lifetime among pixels with no etch pits,³ indicative of spatial nonuniformity in the material that is unrelated to dislocations. (2) An as yet unidentified donor limits the minimum n -type carrier concentrations obtainable during a mercury-saturated low-temperature anneal of the material. While the pressure and temperature dependence of this residual donor does not appear to correlate with the equilibrium dependences of any native point defect,⁶ the nearly universal presence of the donor

in liquid phase epitaxy (LPE), solid-state recrystallized, and molecular beam epitaxy (MBE) materials and its elusive nature do suggest that a native point defect is responsible. (3) Undoped LPE material that has been subjected to a low-temperature mercury-saturated anneal and nominally converted to n type shows an anomalously low mobility. One interpretation is that it is a consequence of interpenetrating p - and n -type regions, with the high effective-mass holes lowering the measured Hall mobilities.⁷ If this model proves to be correct, it may well be a native point defect that causes nonuniform annealing of the material. To overcome this low mobility, a donor impurity is added in concentrations above that of the unknown residual donor. Thus, to lower the n doping to desirable levels, an understanding of the origin of the doping, and the low mobility and a recipe for its elimination are needed. (4) The identity of the primary Shockley-Reed-Hall (SRH) recombination centers has not been established; if they can be correlated with native point defects, strategies for their elimination can be developed. (5) MBE material is often n type as grown⁸ and may be related to a nonequilibrium population of native point defects.

Unraveling the matrix of usually indirect and often contradictory experimental data on the native point defects in semiconductors is a complicated task, especially when the defects may be spatially varying. Numerous theoretical efforts have utilized first-principles methods to elucidate the properties of native point defects in the group IV, III-V, and II-VI semiconductors (see, for example, Refs. 9–12). While these studies have led to much insight into the properties of the point defects, no quantitative predictions of the defect densities were made. Several earlier theoretical studies have looked at the properties of defects in HgCdTe ,^{13–16} although once again no

quantitative predictions of defect densities were given.

Our goal in this paper is to theoretically identify the important native defects in HgCdTe, to calculate their densities as a function of growth and processing conditions, to substantiate the experimentally deduced properties of the native defects, and to begin to unravel the remaining mysteries in this material. To express the concentration of the native point defects in terms of their formation free energies, we employ the quasichemical formalism. In addition to the electron and hole, we have included eight native point defects (and their ionized species) in the analysis: the mercury and tellurium vacancies, the mercury and tellurium antisites, and two types of mercury and tellurium tetrahedral interstitials—one surrounded by four cation near neighbors and one surrounded by four anion near neighbors. As we will show, we have attempted to incorporate all of the important contributions to the free energy and adopt a first-principles approach for most of the quantities we calculate. The only significant empirical data we employ are those needed to obtain the temperature-dependent intrinsic reaction constant. Our calculated native defect concentrations are in quantitative agreement with the available experiment data. Preliminary results of this work have been published previously.¹⁷

A number of features make our study of defects in $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$ unique and permit us to calculate absolute defect concentrations.

(i) To calculate the electronic contribution to the defect formation free energies we employ the self-consistent first-principles full-potential (FP) linearized muffin-tin orbital (LMTO) method¹⁸ and the local-density approximation (LDA). The LMTO method is well suited for compounds containing d electrons, such as $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$.

(ii) Because the LDA overbinds, we have also employed gradient corrections to the LDA of the Langreth-Mehl-Hu type.¹⁹ These corrections greatly improve the overbinding found in the LDA.²⁰ We believe the calculated energies are precise enough that we may make comparison with atoms referenced to the free atom, and therefore by combining these energies with the translational energy of the atoms in the vapor phase, we are able to calculate the chemical potential for a monoatomic mercury vapor.

(iii) A Green's function formalism within a valence force model plus point-charge ionic model is used to calculate the vibrational contribution to the defect-formation free energy, both the enthalpy and entropy.

(iv) The combination of the electronic, translational, and vibrational free energies calculated in (i)–(iii) encompass the primary contributions to the total defect-formation free energies when referenced to a mercury vapor. The total defect-formation free energy is then incorporated into a quasichemical formalism,²¹ and predictions of absolute defect concentrations as a function of the thermodynamic variables temperature and pressure can be made.

The remainder of the paper is organized as follows. In Sec. II we describe the quasichemical formalism used to calculate the neutral native defect concentrations and its extension for ionized defects and alloys. The calculations of the defect-formation free energies are discussed

in Secs. III and IV for the electronic and vibrational contributions, respectively. In Sec. V we present the results of our calculations and a comparison with available experimental results. We end with a brief summary and conclusions of our work in Sec. VI.

II. DEFECT CONCENTRATIONS

A. Quasichemical formalism in compounds

We begin by outlining the formalism employed to calculate the defect concentrations in a compound as a function of external parameters. In Sec. IIC we discuss extensions of the formalism necessary to treat the low- x alloy $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$.

The defect reactions for the compound AC to be considered in this paper are listed in Table I. We have chosen the AC unit cell and A in the phase R as our reference states for the calculation of the reactions' free energies. From Gibbs's phase rule we know that for a system of two components (A and C) and two phases (one of which is the AC zinc blende solid) there are two degrees of freedom. For this paper we shall assume that temperature is one degree and that the chemical potential of an external reservoir of A or C atoms is the other. The chemical potential is chosen to be consistent with the experimental situation to be modeled. We shall choose our reference state to be the mercury vapor and therefore choose to reference our reaction energies to the AC solid and A_R , which will be taken as the monoatomic mercury vapor (extensions of this analysis to the alloys are discussed in Sec. V).

Reference to other reservoirs can be obtained by con-

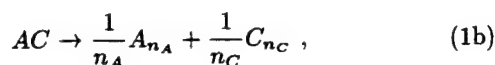
TABLE I. Defect reactions considered for compound AC . The notation is as follows: The primary symbol refers to the species, the subscript refers to the site that the species occupies, with no subscript indicating that the species is occupying its usual lattice site. V corresponds to a vacancy, I an interstitial, R some external reference state, and X a generic defect. Two types of interstitials are considered, both occupying tetrahedral sites, the first surrounded by four cation nearest neighbors I_C , the second surrounded by four anion nearest neighbors I_A . Following the notation of Kröger (Ref. 21), an \times superscript corresponds to a neutral species, a prime to a negatively charged species, a bullet to a positively charged species, and e' and h^\bullet are an electron and a hole, respectively.

AC	\rightarrow	$V_A^\times C + A_R$	(1)
A_R	\rightarrow	AV_C^\times	(2)
$2A_R$	\rightarrow	AA_C^\times	(3)
$2AC$	\rightarrow	$C_A^\times C + 2A_R$	(4)
A_R	\rightarrow	$A_{I_A}^\times$	(5)
A_R	\rightarrow	$A_{I_C}^\times$	(5')
AC	\rightarrow	$C_{I_A}^\times + A_R$	(6)
AC	\rightarrow	$C_{I_C}^\times + A_R$	(6')
X^\times	\rightarrow	$X^{2i} + zh^\bullet$	(7)
X^\times	\rightarrow	$X^{2\bullet} + ze'$	(8)
$e' + h^\bullet$	\rightarrow	0	(9)

sidering the additional reaction



or for molecular species as the reference



where n_i is the appropriate integer, and by taking linear combinations of these with the reactions in Table I.

We are not restricted to specifying the temperature and a chemical potential of one of the species as the two degrees of freedom, but could choose instead another particular set, including, for example, the stoichiometry of the system.¹⁰ While other choices are possible, usually one does not know *a priori* the material's stoichiometry.

Assuming that the defect densities are small and that they are noninteracting, from the law of mass action we can write the reaction constant for each of the neutral defect reactions in Table I as^{21,22}

$$K_{X^\times} \equiv \theta \exp \left(-\frac{F_{X^\times}}{k_B T} \right) = [X^\times], \quad (2)$$

where $[X^\times]$ is the density of the neutral defect X , θ is the number of unit cells per volume and converts $[X^\times]$ from site fraction to defects per unit volume, and F_{X^\times} is the free energy for the neutral defect reaction. The free energy for any defect X can be written as the sum

$$F_X = F_X^{\text{vib}} + F_X^{\text{elect}} + F_X^{\text{trans}} + k_B T \ln(G), \quad (3)$$

where k_B is Boltzmann's constant, and F_X^{vib} is the vibrational, F_X^{elect} is the electronic, and F_X^{trans} is the translational energy contribution to the reaction free energy, and G accounts for the degeneracies of the reactants. In the quasichemical approximation, $G = (g_C g_D)/(g_A g_B)$ for the reaction $A + B \rightarrow C + D$, where g_i is the degeneracy of reactant i . Once F_{X^\times} is known for a given reaction, K_{X^\times} can be evaluated and the defect concentration can be determined. The difficulty, of course, is in the evaluation of F_{X^\times} , which is discussed in Secs. III and IV.

B. Ionized defects and the intrinsic reaction constant

The above discussion applies to the neutral defect densities. In most semiconductors the native point defect will have one or more localized levels in the band gap, allowing for multiple ionization states of the defect. We thus need to calculate the concentration of these ionized defects, in addition to the neutral concentrations discussed above, to obtain the total defect populations. Once the energies of the localized levels are determined (Sec. III E) their populations can be calculated via

$$\frac{[X']}{[X^\times]} = \frac{g_{X'}}{g_{X^\times}} \exp \left(\frac{\mu_F - E_a - F_{X'}^{\text{vib}} + F_{X^\times}^{\text{vib}}}{k_B T} \right) \quad (4)$$

for an acceptor and

$$\frac{[X^*]}{[X^\times]} = \frac{g_{X^*}}{g_{X^\times}} \exp \left(\frac{E_d - \mu_F - F_{X^*}^{\text{vib}} + F_{X^\times}^{\text{vib}}}{k_B T} \right) \quad (5)$$

for a donor state of the defect X . A bullet superscript indicates a positive charge and a prime a negative charge, E_a and E_d are the acceptor and donor one-electron ionization energies with respect to the valence and conduction band (both defined as positive for states in the gap), and μ_F is the Fermi energy. Although the last term in the exponential, $-F_{X'}^{\text{vib}} + F_{X^\times}^{\text{vib}}$, should rightly be there, and corresponds to the difference in the vibrational free energy of the neutral and ionized defect, it has never been considered previously and for the present we shall neglect it too.

For multiply ionized defects with positive Hubbard U 's [reactions (7) and (8) in Table I], the above expression for the number of ionized acceptors generalizes to

$$\frac{[X^{z'}]}{[X^\times]} = \frac{g_{X^{z'}}}{g_{X^\times}} \exp (z\mu_F - E_a^1 - \dots - E_a^z - F_{X^{z'}}^{\text{vib}} + F_{X^\times}^{\text{vib}}), \quad (6)$$

where z is an integer and E_a^i is the ionization energy of the i th ionization level. A similar generalization applies for the donor levels.

For each ionized defect concentration, we introduced one equation. In addition, though, we have two new unknowns, the Fermi energy and either the electron or hole concentration. Thus two additional equations are needed.

First, we have the additional reaction for the generation of electron-hole pairs across the band gap, reaction (9) in Table I, and the corresponding intrinsic reaction constant

$$K_{pn} = [h^*][e'] = pn, \quad (7)$$

where $p \equiv [h^*]$ and $n \equiv [e']$. In general, K_{pn} depends on the shapes of the conduction and valence bands, the band gap energy, the Fermi energy (for degenerate statistics), and the temperature variation of these quantities. Several limits are often encountered in the evaluation of K_{pn} . First, when the conduction and valence bands are parabolic, although not necessarily isotropic, $E \propto k^2$, and the reaction constant can be written in terms of the Fermi-Dirac integrals as

$$K_{pn} = 4 \left(\frac{2\pi k_B T}{h^2} \right)^3 (m_h m_e)^{3/2} \times \mathcal{F}_{1/2} \left(\frac{\mu_F - E_c}{k_B T} \right) \mathcal{F}_{1/2} \left(\frac{E_v - \mu_F}{k_B T} \right), \quad (8)$$

where $\mathcal{F}_{1/2}$ is the Fermi-Dirac function; E_c , E_v , and μ_F are the conduction band, valence band, and Fermi energies, respectively; m_h and m_e are the hole and electron density-of-states effective masses, respectively; and h is Planck's constant. In the nondegenerate limit, this reduces to the familiar expression

$$K_{pn} = 4 \left(\frac{2\pi k_B T}{h^2} \right)^3 (m_h m_e)^{3/2} \exp \left(\frac{E_v - E_c}{k_B T} \right), \quad (9)$$

which is independent of the Fermi energy. For the general case, which will apply even at moderate temperatures for narrow-gap HgCdTe, K_{pn} depends on the extrinsic carrier densities through its dependence on μ_F . Moreover, the conduction band in this narrow-gap material is not well represented by a parabola, but is rather more hyperbolic.²³ Because the hyperbolic dispersion relation approaches a linear dependence away from the band edge, we will for the present assume a linear dispersion relationship of the form $E = \alpha k$. In this case the intrinsic reaction constant becomes

$$K_{pn} = 2 \left(\frac{2\pi k_B T m_h}{h^2} \right)^{3/2} \frac{2}{\pi^2} \left(\frac{k_B T}{\alpha} \right)^3 \times \mathcal{F}_2 \left(\frac{\mu_F - E_c}{k_B T} \right) \mathcal{F}_{1/2} \left(\frac{E_v - \mu_F}{k_B T} \right), \quad (10)$$

where \mathcal{F}_2 is the Fermi-Dirac integral of order 2.

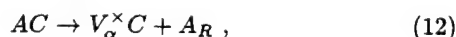
The requirement of charge neutrality leads to a second additional equation:

$$\sum_i \sum_z z [X_i^{z'}] + [e'] = \sum_i \sum_z z [X_i^{z''}] + [h^*], \quad (11)$$

where i sums over the various defects and z sums over the various ionization states of the defect X_i .

C. Quasichemical formalism in alloys

We wish to generalize the above formalism to the ideal cation substituted pseudobinary alloy $A_{1-x}B_xC$. It is perhaps easiest to demonstrate the generalization with a specific defect reaction, for example, a neutral vacancy on the cation sublattice. In the compound AC the formation reaction is given by reaction (1), Table I:



where V_α indicates a vacancy on the cation sublattice. In the alloy a neutral vacancy on the α sublattice can form via the same reaction. The only difference comes in the evaluation of the reaction constant. In the compound the vacancy density is given by

$$[V_\alpha^\times] = \theta \exp \left(\frac{-F_{V_\alpha^\times}}{k_B T} \right) \quad (13)$$

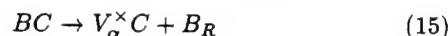
and in the alloy it is given by

$$[V_\alpha^\times] = \theta(1-x) \exp \left(\frac{-F'_{V_\alpha^\times}}{k_B T} \right), \quad (14)$$

where $-F'_{V_\alpha^\times}$ is the reaction free energy corresponding to Eq. (12) in the alloy. The factor of $(1-x)$ results from the configurational entropy contribution to the chemical potential of A on a lattice site (ls), $k_B T \ln([A_{ls}]/[ls]) \simeq k_B T \ln(x)$. Because we have assumed that the defect concentrations are small, the configurational entropy contribution to the chemical potential of A in the compound AC is $k_B T \ln([A_{ls}]/[ls]) \simeq 0$. In both the compound and

the alloy θ is the same.

In addition to Eq. (12), in the alloy the vacancy on the α sublattice can also form via the reaction



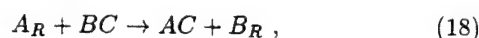
with

$$[V_\alpha^\times] = \theta(x) \exp \left(\frac{-F''_{V_\alpha^\times}}{k_B T} \right), \quad (16)$$

where $F''_{V_\alpha^\times}$ is the reaction free energy corresponding to Eq. (15) in the alloy. Now the vacancy concentrations predicted by Eqs. (14) and (16) must be equal and thus

$$(1-x) \exp \left(\frac{-F'_{V_\alpha^\times}}{k_B T} \right) = (x) \exp \left(\frac{-F''_{V_\alpha^\times}}{k_B T} \right). \quad (17)$$

It is apparent that this simply corresponds to the difference of Eqs. (12) and (15)



that is, the exchange of an A and B on a lattice site, which is a reaction in the alloy in equilibrium with A_R and B_R . An analysis similar to the above applies for the anion antisite C_α , which also substitutes on the cation sublattice.

In the pseudobinary alloy, the vacancy free energies $F_{V_\alpha^\times}$ and $F'_{V_\alpha^\times}$ appearing in Eqs. (13) and (14) may be different. The energy of a vacancy depends on the local configuration of the surrounding lattice; this changes in the second and more distant neighbor shells for the cation vacancy and the anion antisite in the cation substituted alloys. A completely rigorous approach would treat each kind of vacancy uniquely; indeed in a previous work¹⁵ we found a configuration dependence of the vacancy in the $A_{0.5}B_{0.5}C$ lattice of several tenths of an eV, varying approximately linearly in the number of A atoms in the second neighbor shell. Here we have ignored this refinement and assumed the A vacancy and C antisite surroundings are totally of species A . This is justified to some extent because we are interested in low x compositions of $Hg_{1-x}Cd_xTe$.

For defects on the C sublattice (e.g., the anion vacancy and the cation antisites) as well as interstitial atoms in certain tetrahedral sites, one must more carefully consider the complications from the alloy because disorder is found already for nearest neighbors. For systems in which these classes of defects are important, the configuration dependence of the surrounding sublattice must be taken into account. For low x $Hg_{1-x}Cd_xTe$ we find the densities of these defects to be quite low and thus the error incurred in using the electronic energies calculated for the pure AC compound will not impact the major conclusions of our work.

Finally, the band gap is one other important consideration when comparing the alloy to the pure compound. This is of particular importance for the intrinsic reaction constant and we employed an empirical fit to the temperature dependence of the alloy band gap and intrinsic

carrier concentration that are used to calculate K_{pn} . This is presented in detail in Sec. III G.

III. CALCULATION OF DEFECT FORMATION FREE ENERGY: ELECTRONIC CONTRIBUTION

A. Full-potential LMTO calculations

Total energy calculations for the defect reactions were all obtained with a full-potential version²⁴ of the LMTO method in the local-density functional approximation of von Barth and Hedin.²⁵ This method has been tested extensively for most of the elemental *sp* and *d* bonded solids, the II-VI, III-V, and column IV semiconductors, and a host of other solids. Our results, which will be reported elsewhere,²⁰ show good agreement with experiments for all systems studied, with small and systematic errors in structural and mechanical properties. The most prominent error, particularly for the present purposes, is the overbinding of the solid.

In the FP LMTO method, the only important approximation we make beyond the local-density approximation lies in the treatment of the interstitial matrix elements. The LMTO method employs an atom-centered basis, represented by Hankel functions in the interstitial. For the calculations presented here, the basis consisted of a "triple kappa" basis 22 orbitals per atom, with energies -0.01 , -1 , and -2.3 Ry for the *s* and *p* orbitals and -0.01 and -1 Ry for the *d* orbital. Inside the muffin-tin (MT) spheres, wave functions are represented by spherical harmonics and numerically tabulated radial functions. The electron density and potential can be similarly represented since the density generated by a Hamiltonian is obtained by summing over the eigenvectors. Outside the MT spheres, another treatment is necessary. Methfessel²⁶ developed a simple, efficient way to represent the density and potential in the interstitial by extrapolation from the edges of MT spheres, where the value is well known. The electron density is represented in the interstitial as a linear combination of Hankel functions that are chosen to match the value and slope of the function at each MT sphere. Two Hankels per site and *lm* are enough to match the values and slopes at all MT spheres. This representation of the density throughout the interstitial is approximate, although it becomes exact near any MT sphere. Extensive tests show that the approximation works very well for close-packed systems, but the errors can become significant when the packing is poor. To ensure a good fit to the charge density and potential in the interstitial region of the zinc blende solids, we include empty spheres at each tetrahedral interstitial site (rendering the sphere packing bcc for the ideal lattice). In addition, we added orbitals to the basis by centering them on the empty spheres. The addition of 2*s* and 2*p* orbitals changed the energy by approximately 0.1 mRy/atom, showing that the basis is nearly complete.

To assess the validity of the interstitial approximation for the representation of the charge density and interstitial matrix elements, an alternative approach was developed,²⁴ which is similar to a procedure described by

Jones and Sayyesh.²⁷ When calculated in this way the total energies changed by approximately 1 mRy/atom, showing that the approximation is a good one.

Both the charge density inside the spheres and the tails of Hankel functions centered on a neighboring sphere were expanded to $l = 6$. We estimate that the error introduced by truncation at $l = 6$ to be about 1 mRy/atom, in line with other errors in the method. The core was allowed to relax during the self-consistency cycle. The semicore *d* electrons in the tellurium were treated explicitly as valence states in a second panel; explicit treatment of these states was found to introduce a small but significant correction to the total energy. For the 16-atom cells, the Brillouin zone integrals were done by a sampling method for the charge density and the linear tetrahedron method for the band-structure energy, augmented by Blöchl weights, and a mesh of four divisions was used (six *k* points). Checks showed that this was sufficient to converge the energy to 1 mRy/cell.

B. Supercell approximation

Supercells are used in which a periodic array of defects is constructed. Defect formation energies are calculated from a difference in total energies of the compound with and without the defect. For example, if we denote $\mathcal{E}_j(V_A)$ as the energy of a supercell containing *j* lattice sites and one *A* vacancy, the energy for defect reaction (1) in Table I is given by

$$E(V_A) = \mathcal{E}_j(V_A) + E(A_R) - \mathcal{E}_j(AB), \quad (19)$$

where $E(A_R)$ is the energy of an *A* atom in the reference state *R* and $\mathcal{E}_j(AB) = jE(AB)$, where E_{AB} is the energy of an ideal *AB* unit cell. For some defects the number of lattice sites changes in the reaction; for example, for the formation of the *B* antisite via reaction 4 in Table I, the formation energy is given by

$$E(B_A) = \mathcal{E}_j(B_A) + 2E(A_R) - \mathcal{E}_{j+2}(AB). \quad (20)$$

Because we wish to calculate the formation energies in the dilute limit, we use the largest supercell computationally feasible. For this paper, all calculations were done using 16-atom supercells.

C. Gradient corrections to the local density

The local-density approximation generally overbinds the solids. Several systematic extensions of the local-density function have been proposed that are based on generalized gradient approximation for the exchange and correlation energies. We have considered one of these extensions, that proposed by Langreth and Mehl,¹⁹ and have examined the systematics in the gradient corrections to the lattice constants, cohesive energies, bulk modulus, and other elastic constants for a wide array of solids;²⁰ preliminary results of that work for the zinc blende semiconductors are shown in Table II. With few exceptions,

TABLE II. Bulk cohesive energies with and without the gradient corrections (GC) to the local density (LD), and comparison with experiment.

Compound	Cohesive energy (eV/bond)		
	LD	GC	Experiment
Si	2.58	2.31	2.32
Ge	2.22	1.88	1.94
AlP	2.35	2.05	2.13
AlAs	2.17	1.85	1.89
AlSb	1.91	1.61	1.76
GaP	2.08	1.76	1.78
GaAs	1.91	1.56	1.63
GaSb	1.70	1.36	1.48
InP	1.89	1.56	1.74
InAs	1.77	1.42	1.55
InSb	1.60	1.26	1.40
ZnS	1.82	1.53	1.59
ZnSe	1.64	1.35	1.29
ZnTe	1.43	1.15	1.20
CdTe	1.33	1.04	1.10
HgS	1.29	0.94	1.02
HgSe	1.19	0.84	0.85
HgTe	1.09	0.76	0.81

the gradient correction systematically improves the prediction of the cohesive energy, although the prediction of the elastic constants often worsens slightly. The improvement in the cohesive energy is largely due to improvement in the calculation of the total energy of the free atoms, rather than the solid.

Here we are interested in calculating total energies for reactions in which a constituent is exchanged between the solid and the vapor, and thus the errors inherent in the local-density calculation of the cohesive energy will be present in these energies also. Because the relaxations do not change significantly when gradient corrections are added, we have completed the majority of the calculation, including the relaxation, within the FP LMTO. The gradient correction energy, calculated at the LDA-determined relaxed positions, is then added to the LDA energy.

D. Relaxation

In general the lattice relaxes in the presence of a defect, thereby lowering the lattice energy. In the dilute defect limit, the radial relaxation of the lattice extends to infinity.²⁸ In the supercells we account for this relaxation by allowing the overall lattice constant of the supercell to relax to minimize the supercell total energy. Because HgTe and CdTe are nearly lattice matched and their elastic constants are the same, the defect relaxations in pure HgTe should be comparable to those in the HgCdTe alloys. Second, for the most important defects, we permit

the radial relaxation of the defect near-neighbor atoms. For the on-site defects (the vacancies and antisites) we permit only the nearest-neighbor atoms to relax. Estimates of these relaxation energies are given in Sec. V.

Relaxation energies are calculated only for the neutral defects and are assumed comparable in the ionized defects. Nonradial relaxations such as the trigonal and tetragonal distortions that split the degeneracy of the triply degenerate T_2 states may be important and may differ substantially for the different charge states of the system. Because the symmetry of the distortion depends on the charge state of the defect, distortions and charge states must be treated simultaneously. These distortions have not been considered in this paper.

E. Localized defect levels

The calculation of the ionization states of the defects is perhaps the most difficult part of the calculation of the native defect concentrations, in a large part because of the inadequacies of the LDA in predicting the band gap of the semiconductors. This is additionally complicated by the fact that our calculations were done for HgTe, which is known experimentally to be a semimetal with a negative band gap of -0.3 eV, so that even if the LDA band gap were correct, we would still have a zero-gap material. Furthermore, because the Coulomb fields associated with a defect may be extended, we expect that very large supercells will be needed to isolate the localized levels of an individual defect.

We have developed a method to calculate the location of localized defect levels in the band gap and have applied it to the arsenic antisite defect in GaAs. This defect was chosen because of its technological importance and because these levels have been determined experimentally by Weber *et al.*²⁹ Calculations were done within the atomic-spheres approximation so that we could examine the convergence of our results going to large (128-atom) supercells. Our approach is similar to that discussed by Van de Walle *et al.*³⁰ in which the shift in the Fermi level is examined as electrons are added to (or removed from) the defect, with a compensating uniform background charge added so as to maintain charge neutrality. We find good agreement with experiments of the two antisite donor levels. We also have found these energies agreed closely with the positions of peaks in the density of states, when referenced to the top of the valence band. Details of the calculation will be given elsewhere.

Because the compositions of HgCdTe of interest here have narrow band gaps, the determination of the exact location of the defect levels in the band gap is not as important for the purpose of calculating the defect concentrations as in a wider-gap semiconductor such as GaAs, although the identification of the position of defect levels is useful in understanding mechanisms limiting carrier lifetimes. We have used the 54-atom supercells of HgTe to determine the type (acceptor or donor) of the various native defects based on the position of the Fermi level with respect to the states that lie within ~ 0.1 eV above

the valence band edge. In addition, an assessment of whether the state is a single or double donor or acceptor and whether the state is shallow or deep has been made based on the position of the density of states peaks. For the mercury vacancy, we follow the arguments of Cooper and Harrison³¹ and assume that it is a negative- U center, with the neutral and double acceptor states being the only observable states; this assumption is consistent with the observation that the mercury vacancies are always found to be doubly ionized acceptors.¹

F. Ionization state degeneracy

The degeneracy of the various ionization states of each defect may differ and must be determined³² to complete the calculation of the density of ionized defects. As an example, we consider the A vacancy in a II-VI material and use tight-binding language for the purpose of discussing the defect states. There are four dangling anion hybrids, each donating 1.5 electrons to the system, for a total of six electrons at the vacancy site. Although we have not explicitly calculated it, for the purpose of computing state degeneracy, we assume that a symmetry lowering Jahn-Teller distortion will take place whenever there is a state degeneracy beyond two (for spin) and a partial occupancy of that state. Thus we assume that the highest filled vacancy level in the neutral state is doubly occupied with one electron spin up and one spin down, and that the level can accept no other electron. Because there is only one unique configuration for this state, the state has a degeneracy of one. For the single acceptor state in which one electron has been added to the vacancy, the extra electron can either go in spin up or spin down, with equivalent energies. The degeneracy of the state is therefore two. Finally, if the vacancy is a doubly ionized acceptor, the lowest energy configuration for the two additional electrons is with one spin up and one spin down, with a net state degeneracy of one. This assignment of degeneracies—one, two, and one for the neutral, singly ionized, and doubly ionized acceptor, respectively—will hold even if the state is a Hubbard negative- U state, although in this case, the singly ionized state will not be occupied.

A similar argument follows for the other donor and acceptor defect levels. In general, for the II-VI materials we find a degeneracy of one for the neutral defect state, two for the singly ionized state, and one for the doubly ionized state.

G. Intrinsic reaction constant

We are interested in calculating native point defect densities at the relatively high temperatures at which equilibration occurs. It is difficult to calculate K_{pn} theoretically because of the difficulty in calculating the finite-temperature band structure; in general the band gap and the conduction and valence band shapes are all temperature dependent. An additional complexity in calculating K_{pn} in low x $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ is the nonparabolicity of the

conduction and light-hole bands near their extrema.²³

For the purposes of evaluating the defect concentrations, we have calculated K_{pn} using Eq. (10) with $m_h^* = 0.443$,³³ an empirical relationship for the dependency of the band gap on composition x and temperature³⁴

$$\begin{aligned} E_g(x, T) &= E_c - E_v \\ &= -0.302 + 1.93x - 0.810x^2 + 0.832x^3 \\ &\quad + 5.35 \times 10^{-4}T(1 - 2x), \end{aligned} \quad (21)$$

which was fit for $4.2 \leq T \leq 300$ K; and a linear dispersion relationship for the conduction band with α chosen to yield good agreement with experimental values of the intrinsic carrier concentrations³³ for $T < 400$ K. We assume that the intrinsic reaction constant thus computed is valid at temperatures up to 655°C , although there have been no measurements above ~ 400 K to substantiate this extrapolation.³⁵

IV. CALCULATION OF DEFECT-FORMATION FREE ENERGY: VIBRATIONAL CONTRIBUTION

When a defect is introduced into the lattice, the vibrational modes of the system are modified. We must include in our calculation of the defect formation free energy a term that comes from modifications of the vibrational spectrum. Most authors neglect this contribution to the formation free energy. As we will see in Sec. V, although the electronic contribution to the free energy is dominant, the vibrational changes can be significant and they make a substantial impact on the calculated magnitude of the defect concentrations.

Although ideally the vibrational contribution to the formation energy should be calculated within LDA on the same footing as the static electronic contribution, including all of the anharmonic terms, this is a difficult and computationally demanding task. Instead, we take an alternative approach and calculate the vibrational spectrum of the zinc blende lattice using Keating's valence force-field model for the short-range elastic interactions.³⁶ Although experimental elastic constants were used, LDA theory actually predicts the elastic constants within 10% for HgTe and CdTe , so we could equally well have used the calculated values. Because we are dealing with an ionic crystal, we have included a point-charge model to account for the Coulombic interactions.³⁷ Unlike the valence force-field contributions to the dynamical matrix, the Coulomb contributions are long range in nature and induce a macroscopic electromagnetic field, which results in a screening of the transverse optical phonons. The ionic charge is chosen to yield agreement with experiments for the zone center splitting of the transverse and longitudinal optical phonons. A Green's function approach is used to evaluate the lattice-defect-induced modifications to the phonon spectrum; from the perturbed phonon density of states the change in the vibrational free energy can be calculated. Like the electronic energies, the calculations were done for pure HgTe and are assumed applicable to

TABLE III. Neutral native defect formation energies for HgTe corresponding to reactions in Table I, where A is mercury, B is tellurium, and the free atom is used as the reference state A_R . Local-density (LD) calculations were done using a 16-atom supercell, unless otherwise noted. Gradient correction (GC) energies are discussed in the text. The most important ionization states of the native point defects are also given.

Defect	Energy (eV)				Ionization state
	LD	Relaxation	GC	Total	
V_{Hg}	2.83	-0.05	-0.69	2.09	double acceptor ^a
V_{Te}	0.99	-0.01	0.49	1.47	shallow donor
Hg_{Te}	-1.00	0	1.46	0.46	deep acceptor
Te_{Hg}	4.85	-0.19	-1.54	3.12	shallow donor
$Hg_{I_{Hg}}$	0.75	-0.24	0.70	1.21	shallow donor
$Hg_{I_{Te}}$	0.81	-0.31	0.62	1.12	shallow donor
$Te_{I_{Hg}}$	4.78	-0.57	-0.83	3.38	shallow donor
$Te_{I_{Te}}$	5.17	-0.84	-0.96	3.37	shallow donor

^aAssumed to be a negative- U center.

defect calculations in $Hg_{0.8}Cd_{0.2}Te$. We expect that this is a reasonable assumption because the elastic constants for HgTe and CdTe are nearly identical. In this paper we only consider the vibrational free energy of the neutral defects and assume

$$F_{X^{vib}}^{vib} \simeq F_{X^{**}}^{vib} \simeq F_{X^{\times}}^{vib} \quad (22)$$

for all ionization states. Details of the calculations are given in the Appendix. Preliminary results of this model were given previously.^{14,17}

V. RESULTS AND DISCUSSION

A. Formation free energies in HgTe

Calculated defect-formation electronic energies in HgTe for the defect reactions listed in Table I and with A_R as the free mercury atom are listed in Table III. For all defects, the gradient correction for the 16-atom supercell is calculated for the relaxed configuration, as determined by the LDA calculation.

Although the total electronic formation energies listed in Table III are important contributions to the formation free energy, these energies alone cannot be used to assess the relative importance of the various defects in the solid. This is mostly due to the free energy of the atom in the reference state (in excess of its free atom electronic energy) that is not included in these electronic energies, and which is discussed in the Sec. VB. This point should be obvious because we could have just as well defined our defect reactions with respect to the tellurium molecule in the vapor phase and the HgTe solid and obtained the corresponding reaction energies that would be quite different from those in Table III.

The calculated phonon dispersion curve for HgTe is given in Fig. 1 and is in fair agreement with the experimental results. The discrepancies with the experimental curves, in particular near the Brillouin zone boundary, can be attributed for the most part to our neglect of

long-range elastic interactions in the near-neighbor valence force-field model.³⁸ The vibrational entropy and energy contribution to the defect formation free energy are calculated from the density of phonon states; results at 500°C are given in Table IV. Equation (A20) can be used to estimate the values at other (high) temperatures.

B. Defect reaction constants in $Hg_{0.8}Cd_{0.2}Te$

The reaction constants for each of the defect reactions listed in Table I are calculated as a function of temperature, where the reference state was taken as the monoatomic mercury vapor at pressure P_{Hg} . Electronic energies for the neutral defect formation energies are taken from Table III. Vibrational free energies are calculated using the general temperature expression, as discussed in the Appendix, although for the purpose of obtaining an analytical expression for the reaction constants with the primary temperature dependency explicitly displayed, we have fit our results for 500°C to the high-temperature expression, Eq. (A20).

A third contribution to the formation free energy comes from the free energy of the mercury in the reference state, less the electronic energy of the free mercury atom that is contained in the electronic defect formation

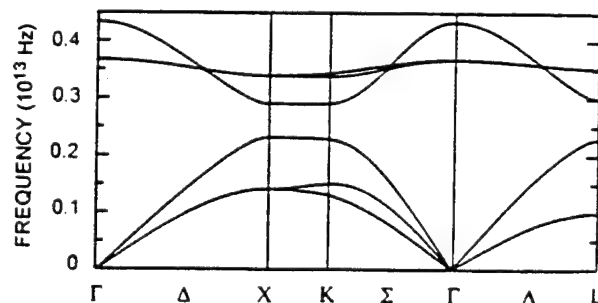


FIG. 1. Calculated phonon dispersion curve for HgTe.

TABLE IV. Entropy (S^{vib}) and energy (U^{vib}) contributions to the vibrational free energy in HgTe at 500 °C, for the defect reactions listed in Table I.

Defect	$S^{\text{vib}}(k_B)$	$U^{\text{vib}}(\text{eV})$
V_{Hg}	-8.9	-0.20
V_{Te}	9.7	0.19
Hg_{Te}	18.7	0.39
Te_{Hg}	-18.7	-0.39
Hg_I	9.5	0.20
Te_I	-9.5	0.20

energies. For the vapor, we must include the translation free energy of the atoms. The chemical potential for monoatomic mercury in the gas phase is given by the standard expression from statistical mechanics

$$\mu_{\text{Hg}} = kT \ln \left[\frac{P_{\text{Hg}}}{kT} \left(\frac{h^2}{2\pi m_{\text{Hg}} kT} \right)^{3/2} \right]. \quad (23)$$

Combining these three contributions to the reaction free energy we obtain the reaction constants for each of the defect reactions. For example, for the neutral mercury vacancy in HgTe, we obtain

$$K_{V_{\text{Hg}}} = [V_{\text{Hg}}^{\times}] = \theta(1-x) \frac{kT}{P_{\text{Hg}}} \left(\frac{2\pi m_{\text{Hg}} kT}{h^2} \right)^{3/2} \times (1.27 \times 10^6 T^{-3}) \exp \left(\frac{-2.09 \text{ eV}}{kT} \right), \quad (24)$$

where we have taken $g_{V_{\text{Hg}}^{\times}} = 1$, $-F_{V_{\text{Hg}}^{\times}}^{\text{elect}} = -2.09$, $\theta = 1.48 \times 10^{22} \text{ cm}^{-3}$, and replaced $\exp(-F^{\text{vib}}/k_B T)$ by $\exp[-(-0.2 \text{ eV} + 8.9 k_B T)/k_B T]_{500^\circ\text{C}} \simeq 1.27 \times 10^6 T^{-3}$. Reaction constants for the eight native point defects considered in this paper are listed in Table V.

TABLE V. Reaction constants for neutral defects in $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$, corresponding to the defect reactions in Table I. The vibrational contribution is calculated at 500 °C and fit to the high-temperature power-law dependence [Eq. (A19)] so as to show the explicit temperature dependence. All calculations in the paper were done using reaction constants with the more exact expression for the vibrational free energies.

Defect	Reaction constant
V_{Hg}^{\times}	$K_{V_{\text{Hg}}^{\times}} = [V_{\text{Hg}}^{\times}] = 1.14 \times 10^{30} T^{-3/2} P_{\text{Hg}}^{-1} \exp(-\frac{2.09}{k_B T})$
V_{Te}^{\times}	$K_{V_{\text{Te}}^{\times}} = [V_{\text{Te}}^{\times}] = 4.55 \times 10^{14} T^{1/2} P_{\text{Hg}} \exp(-\frac{1.47}{k_B T})$
$\text{Hg}_{\text{Te}}^{\times}$	$K_{\text{Hg}_{\text{Te}}^{\times}} = [\text{Hg}_{\text{Te}}^{\times}] = 4.47 \times 10^6 T^1 P_{\text{Hg}}^2 \exp(-\frac{0.46}{k_B T})$
$\text{Te}_{\text{Hg}}^{\times}$	$K_{\text{Te}_{\text{Hg}}^{\times}} = [\text{Te}_{\text{Hg}}^{\times}] = 3.80 \times 10^{37} T^{-1} P_{\text{Hg}}^{-2} \exp(-\frac{3.12}{k_B T})$
$\text{Hg}_{I_{\text{Hg}}}^{\times}$	$K_{\text{Hg}_{I_{\text{Hg}}}^{\times}} = [\text{Hg}_{I_{\text{Hg}}}^{\times}] = 3.03 \times 10^{14} T^{1/2} P_{\text{Hg}} \exp(-\frac{1.21}{k_B T})$
$\text{Hg}_{I_{\text{Te}}}^{\times}$	$K_{\text{Hg}_{I_{\text{Te}}}^{\times}} = [\text{Hg}_{I_{\text{Te}}}^{\times}] = 3.03 \times 10^{14} T^{1/2} P_{\text{Hg}} \exp(-\frac{1.12}{k_B T})$
$\text{Te}_{I_{\text{Hg}}}^{\times}$	$K_{\text{Te}_{I_{\text{Hg}}}^{\times}} = [\text{Te}_{I_{\text{Hg}}}^{\times}] = 7.23 \times 10^{29} T^{-1/2} P_{\text{Hg}}^{-1} \exp(-\frac{3.38}{k_B T})$
$\text{Te}_{I_{\text{Te}}}^{\times}$	$K_{\text{Te}_{I_{\text{Te}}}^{\times}} = [\text{Te}_{I_{\text{Te}}}^{\times}] = 7.23 \times 10^{29} T^{-1/2} P_{\text{Hg}}^{-1} \exp(-\frac{3.37}{k_B T})$

The reaction constants for the ionized defects are calculated using Eq. (6) for acceptors and its generalization for donors. We define

$$K_{X^{z'}} = [X^{z'}] = [X^{\times}] \frac{g_{X^{z'}}}{g_{X^{\times}}} \exp(z\mu_F - E_a^1 - \dots - E_a^z) \\ = K_{X^{\times}} \frac{g_{X^{z'}}}{g_{X^{\times}}} \exp(z\mu_F - E_a^1 - \dots - E_a^z) \quad (25)$$

and

$$K_{X^{z*}} = [X^{z*}] = K_{X^{\times}} \frac{g_{X^{z*}}}{g_{X^{\times}}} \exp(E_d^1 + \dots + E_d^z - z\mu_F). \quad (26)$$

C. Defect concentrations in $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$

Gibbs's phase rule tells us that for a system of three components (A, B, and C) and two phases (zinc blende solid and vapor) there are three degrees of freedom. In evaluating the defect concentrations in $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$ we have chosen the temperature, the mercury pressure P_{Hg} , and the alloy composition x as these specified variables; the tellurium and cadmium pressures, the crystal stoichiometry, and the density of the various native point defects are determined by these conditions.

The reaction constants in Table V are evaluated to determine the concentrations of the various native point defects as a function of temperature and pressure. The Fermi energy is determined by requiring charge neutrality. The activation energies for the shallow donor and acceptor states are taken to be zero; the sensitivity of our results to this assumption is discussed further below.

Figure 2(a)–2(c) show the defect concentrations at various equilibration temperatures. Pressure ranges are chosen so as to stay within the stability region of the material.⁶ At all temperatures and pressures considered, the dominant defect is found to be the doubly ionized mercury vacancy, in agreement with previous interpretations of experiments;¹ our result confirms the generally accepted experimental observation that the mercury vacancy is responsible for the p -type behavior of undoped HgCdTe equilibrated at high temperatures.

At all temperatures, the second most dominant defect is found to be the tellurium antisite. The antisite concentration decreases more rapidly with P_{Hg} than does the mercury vacancy, and thus is most important at low mercury pressures. As does the mercury vacancy, the tellurium antisite defect accommodates excess tellurium in the lattice, and therefore its presence also shifts the stoichiometry towards the tellurium-rich side of the phase diagram.

The reason the tellurium antisite concentrations is so high deserves comment. The tellurium antisite formation energy is larger than that for the mercury vacancy by 1 eV (Table III). However, as can be seen from Table V, the pre-exponential factor of the reaction constant for the tellurium antisite is enormous. The large pre-exponential factor results from the large phase space factor (entropy) gained by creating two free mercury atoms compared to that lost by elimination of a formula unit. In contrast, the mercury antisite density is low, despite the fact that

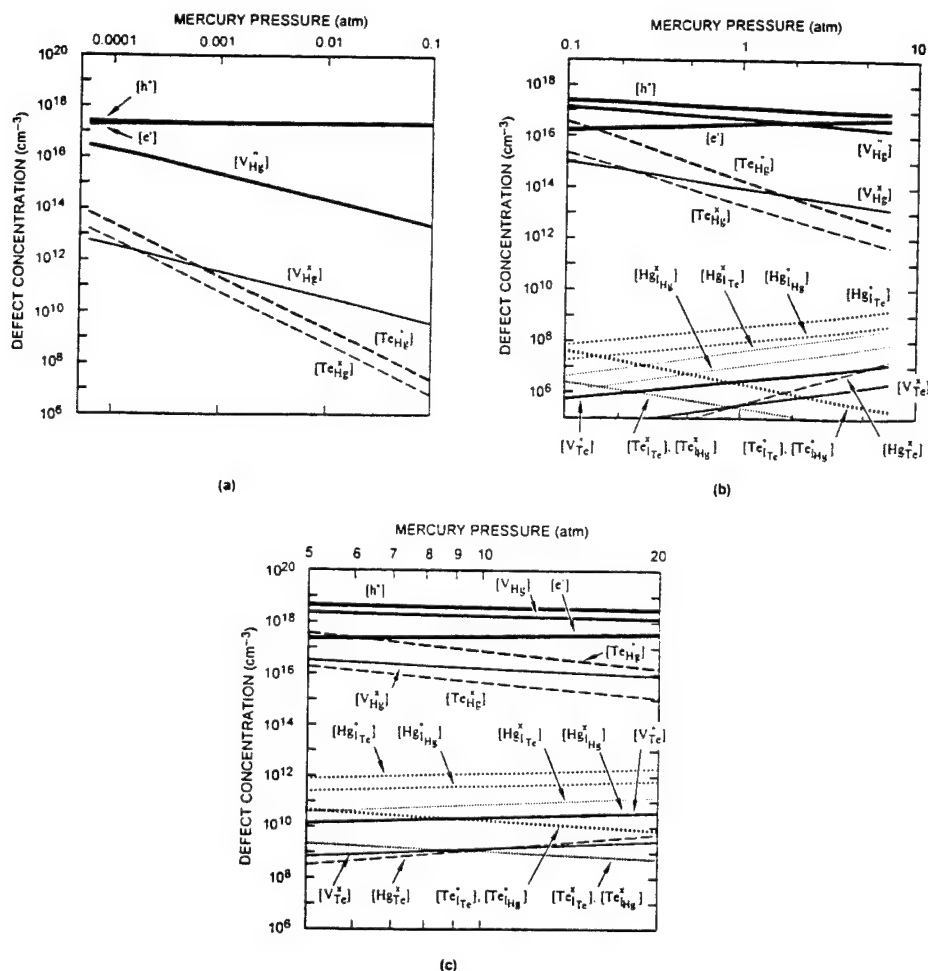


FIG. 2. Defect concentrations at annealing temperature for material annealed at (a) 250°C, (b) 500°C, and (c) 655°C.

its formation energy is quite small; this again is a result of the pre-exponential factor that in this case is very small. Thus it is clear that one must be cautious in deducing the relative populations of the various defects based on the electronic contributions to the defect formation energies alone.

While the tellurium antisite is never found to be the dominant defect controlling the doping under equilibrium conditions, it can introduce significant compensation at low mercury pressures. Moreover, the diffusion coefficient of the antisite is expected to be quite small because the diffusion of an antisite will necessarily involve at least one additional point defect, such as the mercury vacancy or the tellurium interstitial. Thus the tellurium antisite may not reach equilibrium densities for the times and temperatures corresponding to the low-temperature ($\sim 250^\circ\text{C}$), high-mercury-pressure anneals typically employed to reduce the mercury vacancy density. If tellurium antisite densities are in fact equilibrated at a temperature at which the antisite diffusion effectively stops during cool down from the growth temperature, then the antisites may be frozen in at higher, nonequilibrium concentrations. If such a freezing in of nonequilibrium antisites does occur, the tellurium antisite may in fact be the "universal" residual donor observed in densities of $\sim 10^{14}\text{ cm}^{-3}$ in material subjected to a low-mercury-pressure,

high-temperature anneal, although it is not clear why, for example, the frozen-in density of antisites would be the same for LPE material grown from both mercury- and tellurium-rich melts. The tellurium antisite may also be the defect responsible for the n -type carrier concentrations in as-grown MBE material,⁸ which is believed to be grown on the tellurium-rich side of the phase diagram where antisite populations are highest.

Annealing strategies for reduction of the tellurium antisite densities can be developed and may be important if the antisite is the residual donor. Consider, for example, a two-temperature annealing process in which a first anneal is done under mercury-saturated conditions, but at the lowest temperature for which the antisite is able to equilibrate in reasonable times. This anneal would serve to lower the antisite densities as much as possible. A second anneal would be much like that currently employed, that is, at $\sim 250^\circ\text{C}$ under mercury-saturated conditions and would serve to anneal out the mercury vacancies, leaving the antisite densities effectively unchanged.

At 500°C mercury interstitials are present at levels $\sim 10^{10}\text{ cm}^{-3}$ and at no temperatures are present at levels high enough to significantly compensate the mercury vacancies, much less to turn the material n -type under equilibrium conditions. Unlike the tellurium antisites, the mercury interstitials are relatively fast diffusers,² and

thus it is unlikely that nonequilibrium densities of interstitials will be frozen in. The mercury interstitial densities that we predict are in quantitative agreement with those needed to explain the diffusion in $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ in the process simulator developed by Meléndez and Helms.³⁹ There is some uncertainty in the quantitative predictions of the mercury interstitial densities reported here because of the neglect of the alloy effects that we expect to be more significant than for the mercury vacancy and tellurium antisite. Although this correction will be largest for the mercury interstitial surrounded by four mercury first neighbors, it should also be significant for the interstitial surrounded by four tellurium first neighbors because of the six cation second-nearest neighbors, which are only slightly more distant than the first neighbors.

The mercury antisite and the tellurium vacancy and interstitial densities are all quite low, never exceeding $\sim 10^8 \text{ cm}^{-3}$ at 500°C . The corrections to these predicted densities may be sizable because of alloy effects, but such corrections should not significantly impact the densities of the mercury vacancy and tellurium antisite.

Figure 2 shows the defect concentrations at the temperatures at which equilibration takes place; in Fig. 3 we show the defect concentrations for material equilibrated at 500°C , then quench cooled to 77 K. We have assumed that the total defect concentrations are frozen in during the quench, for example, $[V_{\text{Hg}}]_{\text{total}} = [V_{\text{Hg}}^x] + [V_{\text{Hg}}'] + [V_{\text{Hg}}'']$ is constant, but that the electrons and holes are allowed to reach a new equilibrium corresponding to the low temperature. Figure 4 shows the low-temperature hole concentrations for such quench-cooled materials as a function of P_{Hg} and T_{anneal} , compared with the experimental results of Vydyanath.¹ The agreement of our theoretical results with the experiments is remarkably good considering that our calculated results are obtained al-

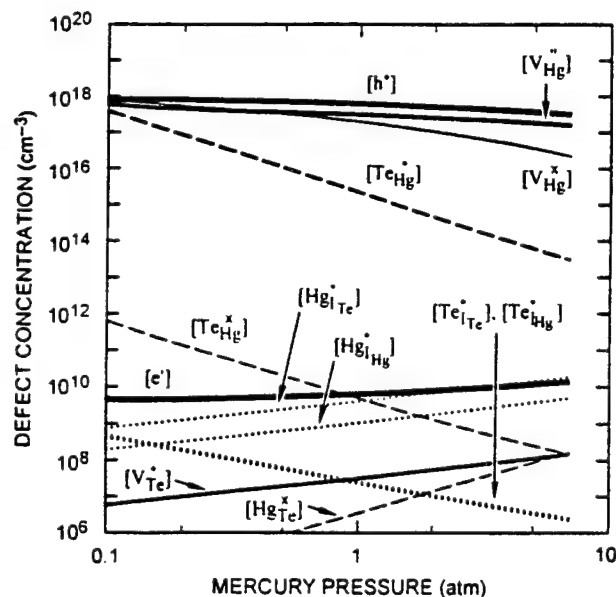


FIG. 3. Defect concentrations for the 500°C anneal after quench cooling to 77 K.

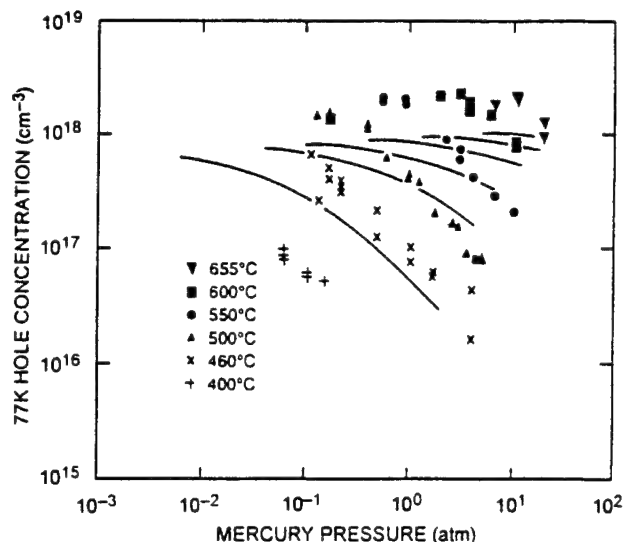


FIG. 4. Hole concentrations at 77 K for material equilibrated at various high-temperature annealing conditions. Experimental results taken from Ref. 1.

most entirely from first principles. Moreover, there is a significant uncertainty in the quenching efficiency of the experiments; thus exact agreement with the experiments is not a valid criterion for testing the accuracy of the theory.

From Fig. 4 one can see that for the higher annealing temperatures our calculations predict a different dependence of hole concentration on the mercury pressure than is experimentally observed. The lower slopes for the theory result from our finding that the material is extrinsic at the higher temperatures. If the material is intrinsic when equilibrated as the experiments indicate, then $[V_{\text{Hg}}'']_{\text{ht}} \propto P_{\text{Hg}}^{-1}$, as can be obtained from Table V, with μ_F independent of $[V_{\text{Hg}}]$. However, if the material is extrinsic when equilibrated, that is, with $[h^*] = 2[V_{\text{Hg}}'']$, then from the reaction



we see that $[V_{\text{Hg}}''] [h^*]^2 \propto [V_{\text{Hg}}'']^3 \propto P_{\text{Hg}}^{-1}$.

The discrepancies between theory and experiments may be due to a number of factors. First, both uncertainty in the quenching efficiencies and analysis of the Hall data may account for some of the discrepancy. There are also a number of uncertainties in the theoretical calculation that may account for the discrepancies. These include uncertainties in the electronic and vibrational defect formation free energy, the ionization energies of the defects (which were assumed to be zero in the above calculations), alloy effects, and finally the uncertainties in the intrinsic reaction constant. These are discussed in turn below.

First it is interesting to examine the sensitivity of our predictions to the accuracy of electronic and vibrational defect formation free energy. In Fig. 5 we have recalculated the 77 K hole concentrations as a function of the annealing temperature with the electronic contribution

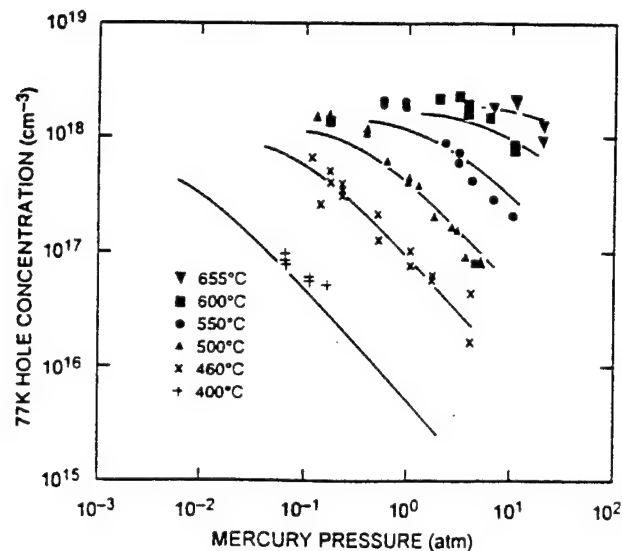


FIG. 5. Hole concentrations at 77 K for material equilibrated at various high-temperature annealing conditions, calculated with the electronic formation energy for the mercury vacancy increased by 10% and a rigid upward shift of the hole concentrations by a factor of 2.2. This figure is meant to demonstrate the sensitivity of our results to small changes in our calculated parameters. As discussed in the text, similar qualitative changes in our results can be seen by modifying the intrinsic reaction constant.

to the mercury vacancy formation energy increased by 10% and a rigid shift upward in our results by a factor of 2.5; such small changes result in better agreement with experiments. The corrections to our calculated mercury vacancy and tellurium antisite formation energies due to alloy effects have not yet been included in these calculations and are expected to be in the range of tenths of an eV, as discussed in Sec. II C, and thus may account for the magnitude of correction used in this example. Preliminary estimates for the corrections for going from the 16-atom to the 32-atom supercell are ~ 0.1 eV for the cation vacancy and the tellurium antisite as well. Finally, Jahn-Teller relaxation energies have not been included in the present work and they may modify the electronic formation free energies. An increase in the effective vibrational frequencies [$\bar{\omega}$ in Eq. (A20)] can account for an upward shift in the densities. Such an increase may arise from differences between the neutral and ionized defect vibrational free energies and, perhaps, anharmonic effects that may be large at defects such as the vacancy where an atom is missing from the lattice. Thus we see that our calculations agree with the experimental data approximately to within the known uncertainties of the theory.

Our results are also very sensitive to the intrinsic reaction constant, which in turn depends sensitively on the band structure and its temperature dependence. HgCdTe is known to be anomalous in that its band gap is found to increase with temperature at room temperature and below, and although there is no experimental information on the temperature dependence of the gap at higher

temperatures, we have assumed that Eq. (21) extrapolates to higher temperatures. In addition, as discussed above, we have assumed a parabolic valence band, but a linear variation of the conduction band, with the slope chosen to agree with the intrinsic carrier concentrations [with the energy gap given by Eq. (21) at temperatures below 400 °C]. While this fit is quite good for the temperature range over which it is fit (from 77 K to 400 K), the reliability of K_{pn} at 250 °C and above for which we have presented our defect density predictions is unknown. To demonstrate the sensitivity of our results to the intrinsic reaction constant we have calculated the 77 K temperature hole concentrations with the conduction band density of states increased a factor of 10 and have found, except at the very highest temperatures, that the calculated hole concentrations vary as P_{Hg}^{-1} , indicating intrinsic behavior at the annealing temperature and resulting in better agreement with experiments. Because our results depend sensitively on the intrinsic reaction constant, it is essential to establish a reliable prediction of its value at the annealing and growth temperatures where equilibration of the defect densities takes place.³⁵

We have suggested that nonequilibrium densities of tellurium antisites may be the residual donor, but they may also be important SRH recombination centers. It is experimentally observed that the residual donor does not freeze out even for samples cooled to 4 K, and therefore its first ionization state must resonate in the conduction band. A SRH recombination center in $Hg_{1-x}Cd_xTe$ with $x = 0.22$ lies ~ 25 meV below the conduction band edge, has a larger capture cross section for electrons than holes, and typically has a density smaller than, but comparable to, the residual donor density. The properties of the antisite are consistent with such a level: it is a donor; although we have assumed here it is a single donor, it is likely that a second donor level is present in the gap and may be ionized at the high processing temperatures; and the first ionization level may be resonant in the conduction band, although we are unable to resolve this in our present calculations with certainty. A more quantitative prediction of the ionization levels of the antisite is needed to correlate it with a SRH center.

A technologically important step in making ir detectors from $Hg_{1-x}Cd_xTe$ is a low-temperature mercury-saturated anneal that is done to reduce the mercury vacancy concentrations. In Fig. 6 we show the defect concentrations for material annealed at various temperatures along the mercury-saturated side of the phase diagram. In such mercury-saturated anneals, if equilibrium can truly be reached, then the mercury vacancies will certainly be the dominant defect, with the tellurium antisite density being negligible. However, as discussed above, it is unlikely that equilibrium densities of antisites will be achieved at these relatively low temperatures.

The above analysis of defect concentrations can be repeated for a number of different situations. For example, we can calculate the native defect densities with a donor or acceptor impurity present. At the high growth temperatures, the impurity concentrations would have to be comparable to the vacancy concentrations to modify the high-temperature vacancy concentration. We can also re-

peat the above calculations for reference state other than the mercury vapor. For example, in HgCdTe tellurium precipitates are known to form as a metastable state upon cooling from high growth temperatures.⁴⁰ In the vicinity of a precipitate, the native defect populations will be in local equilibrium with the tellurium solid and defect concentrations for this reference state can be calculated. Because this constitutes a nonequilibrium situation, one must address diffusion rates to assess the extent of the modified defect atmosphere about a precipitate.

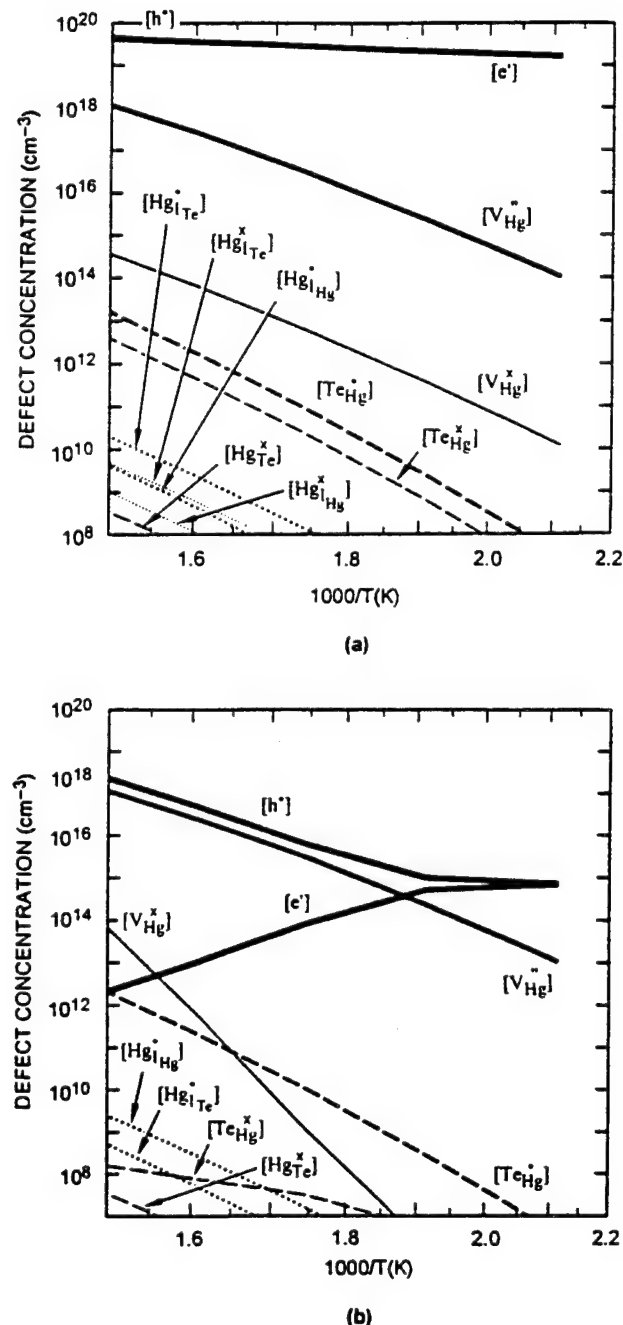


FIG. 6. Defect concentrations for material annealed along the mercury-rich side of the stability region at (a) the annealing temperature and (b) after quenching to 77 K.

VI. SUMMARY AND CONCLUSIONS

We have made quantitative predictions of the native point defect densities in Hg_{0.8}Cd_{0.2}Te as a function of temperature and pressure and find good agreement with the available experiments. We have substantiated the claim that the primary defect is the mercury vacancy and have identified the tellurium antisite as an important secondary defect. A first-principles approach was used for most of the quantities calculated, with the only significant empirical data being those needed to obtain the temperature-dependent intrinsic reaction constant.

Although we predict the undoped material to be always *p* type, refinements in our calculations may show that the antisite may dominate in the low-mercury-pressure region and turn the material *n* type by a native defect; our current accuracy is not sufficient to establish this. While most anneals of technological importance are done under mercury-saturated conditions to reduce mercury vacancy concentrations, exploration of the tellurium-saturated region where we predict the tellurium antisite densities become comparable to those of those of the mercury vacancy may help confirm the presence of tellurium antisites.

A second means to explore the presence of tellurium antisites and their relationship to the residual donor is through a careful set of experiments using two temperature anneals, as discussed above. Because we do not know the temperature at which diffusion of tellurium effectively stops, the temperature of the first anneal would have to be varied, as would the annealing time; the mercury pressure could also be varied, although mercury-saturated conditions are those one would eventually want to employ. The identification of the tellurium antisite as the residual donor can be made if the donor densities in identically grown material were found to differ after the second anneal (using the standard conditions for a mercury-saturated low-temperature anneal) depending on the conditions of the first anneal. A quantitative analysis of this experiment would be quite difficult because, in addition to uncertainties in the temperature at which the tellurium antisite equilibration stops, if the tellurium antisite diffuses via a vacancy mechanism, the diffusion of the antisite will depend on the concentration of mercury vacancies present during the first anneal.

It would also be useful to perform high-temperature annealing as was done by Vydyanath,¹ but instead of quenching to 77 K and having to address the issue of quenching efficiency, follow the anneals by Hall analysis at the anneal temperature. Recently, an attempt at such an experiment was made by Wienecke et al.,³⁵ although an analysis of such an experiment requires knowledge of the high-temperature intrinsic reaction constant.

ACKNOWLEDGMENTS

We wish to acknowledge the work of Michael Methfessel in development of the first-principles codes applied in this work. This work has been supported by NASA Contract No. NAS1-18226, ARPA Contract No. MDA972-

92-C-0053, and ONR Contract No. N00014-89-K-132. Computational support was provided by the Numerical Aerodynamical Simulation computing facility at NASA Ames Research Center.

APPENDIX: VIBRATIONAL ENTROPY OF POINT DEFECTS IN SEMICONDUCTORS

As discussed by Keating,³⁶ we assume the elastic energy for a zinc blende structure can be written as

$$V = \frac{3\alpha}{8d_0^2} \sum_i [\Delta(\vec{r}_i \vec{r}_i)]^2 + \frac{3\beta}{8d_0^2} \sum_{i>i'} [\Delta(\vec{r}_i \vec{r}_{i'})]^2, \quad (A1)$$

where i and i' sum over all bonds, $\Delta(\vec{r}_i \vec{r}_{i'}) = \vec{r}_i \vec{r}_{i'} - \vec{r}_i^0 \vec{r}_{i'}^0$, with \vec{r} and \vec{r}^0 the bond vectors connecting adjacent atoms in the distorted and equilibrium lattices, respectively. For first-neighbor interactions, the sum in the second

term runs over only those bonds connected to a common atom.

For a nonionic material, the normal modes for the displacement of the atoms are determined by solving the equation of motion for the lattice cast in the usual manner in terms of the three-dimensional eigenvalue equation

$$\omega^2 \vec{u} = \mathbf{D}^e(\mathbf{k}) \vec{u}. \quad (A2)$$

Here $\mathbf{D}^e(\mathbf{k})$ is the (elastic) dynamical matrix and

$$\vec{u} = \begin{pmatrix} \vec{u}_1 \\ \vec{u}_2 \end{pmatrix} \quad (A3)$$

is the polarization vector of the normal modes, where \vec{u}_i is the displacement vector of the i th atom. For the zinc blende lattice there are two atoms per unit cell, so $i = 1$ or 2.

In terms of the elastic constants, C_{11} and C_{12} , the dynamical matrix for the valence force field model is given by

$$\mathbf{D}^e(\vec{k}) = \begin{pmatrix} \frac{8dC_{11}}{\sqrt{3}m_1} \mathbf{I} & \frac{2dC_{12}}{\sqrt{3}m_1 m_2} \mathbf{S}(\vec{k}) + \frac{2d(C_{11}-C_{12})s(\vec{k})}{\sqrt{3}m_1 m_2} \mathbf{I} \\ \frac{2dC_{12}}{\sqrt{3}m_1 m_2} \mathbf{S}^*(\vec{k}) + \frac{2d(C_{11}-C_{12})s^*(\vec{k})}{\sqrt{3}m_1 m_2} \mathbf{I} & \frac{8dC_{11}}{\sqrt{3}m_2} \mathbf{I} \end{pmatrix}, \quad (A4)$$

where d is the equilibrium bond length, m_1 and m_2 are the masses of the two atoms in the unit cell, \mathbf{I} is the 3×3 unit matrix, and \mathbf{S} is given by

$$\mathbf{S} = \begin{pmatrix} s_1 & s_4 & s_3 \\ s_4 & s_1 & s_2 \\ s_3 & s_2 & s_1 \end{pmatrix} \quad (A5)$$

with

$$s_1(\vec{k}) = e^{i\vec{k} \cdot \vec{d}_1} + e^{i\vec{k} \cdot \vec{d}_2} + e^{i\vec{k} \cdot \vec{d}_3} + e^{i\vec{k} \cdot \vec{d}_4}, \quad (A6a)$$

$$s_2(\vec{k}) = e^{i\vec{k} \cdot \vec{d}_1} + e^{i\vec{k} \cdot \vec{d}_3} - e^{i\vec{k} \cdot \vec{d}_2} - e^{i\vec{k} \cdot \vec{d}_4}, \quad (A6b)$$

$$s_3(\vec{k}) = e^{i\vec{k} \cdot \vec{d}_1} - e^{i\vec{k} \cdot \vec{d}_2} + e^{i\vec{k} \cdot \vec{d}_3} - e^{i\vec{k} \cdot \vec{d}_4}, \quad (A6c)$$

and

$$s_4(\vec{k}) = e^{i\vec{k} \cdot \vec{d}_1} - e^{i\vec{k} \cdot \vec{d}_2} - e^{i\vec{k} \cdot \vec{d}_3} + e^{i\vec{k} \cdot \vec{d}_4}. \quad (A6d)$$

The \vec{d}_j are the vectors connecting atom 1 to atom 2 in the unit cell and are given by $d_1 = \frac{a}{4}[111]$, $d_2 = \frac{a}{4}[1\bar{1}\bar{1}]$, $d_3 = \frac{a}{4}[\bar{1}1\bar{1}]$, and $d_4 = \frac{a}{4}[\bar{1}\bar{1}1]$, where a is the lattice constant.

While short-range elastic forces are described within the valence force field model, in crystal with an ionic contribution to the bonding, Coulomb interaction must also be included in the dynamical matrix. The long-range nature of the Coulomb interaction complicates the problem considerably. Using a pairwise point-charge model of the Coulomb interaction, the Coulomb dynamical matrix is given by³⁷

$$D_{\alpha\alpha'}^c(\kappa, \kappa' | \vec{k}) = -\frac{\delta_{\kappa, \kappa'}}{m_{\kappa}} \sum_{\kappa''} \phi_{\alpha\alpha'}(\kappa \kappa'' | 0) + \frac{1}{\sqrt{m_{\kappa} m_{\kappa'}}} \phi_{\alpha\alpha'}(\kappa \kappa' | \vec{k}), \quad (A7)$$

where

$$\begin{aligned} \phi_{\alpha\alpha'}(\kappa, \kappa' | \vec{k}) = & -\frac{q_{\kappa} q_{\kappa'}}{4\pi\epsilon_0} P^{3/2} \times \sum_{l'} H_{\alpha\alpha'} \{ \sqrt{P} [\vec{x}(l, \kappa) - \vec{x}(l', \kappa')] \} e^{i\vec{k} \cdot [\vec{x}(l, \kappa) - \vec{x}(l', \kappa')]} \\ & + \frac{q_{\kappa} q_{\kappa'}}{v_0 \epsilon_0} \sum_{\vec{G}} \frac{(\vec{G} + \vec{k})_{\alpha} (\vec{G} + \vec{k})_{\alpha'}}{|\vec{G} + \vec{k}|^2} \exp \left(\frac{-|\vec{G} + \vec{k}|^2}{4P} \right) \exp \{ i\vec{G} \cdot [\vec{x}(\kappa) - \vec{x}(\kappa')] \}. \end{aligned} \quad (A8)$$

In the above equation l and κ label the unit cell and basis atoms, α refers to the Cartesian component, q_{κ} is the effective charge, $\vec{x}(\kappa, l) = \vec{x}(\kappa) + \vec{x}(l)$ is the position vector of the κ th atom in the l th unit cell, \vec{G} are

the reciprocal lattice vectors, v_0 is the unit cell volume, ϵ_0 is the permittivity constant, and P is a (numerically determined) measure of the Gaussian charge distribution used in the Ewald summation. $H_{\alpha, \alpha'}(y)$ is given by the

integro-differential expression

$$H_{\alpha,\alpha'}(y) = \frac{2}{\sqrt{\pi}} \frac{\partial}{\partial y_\alpha} \frac{\partial}{\partial y_\beta} \left(\frac{1}{y} \int_y^\infty \exp(-x^2) dx \right). \quad (\text{A9})$$

The full dynamical matrix is given by the sum of the Coulomb and elastic contributions

$$\mathbf{D}(\vec{k}) = \mathbf{D}^c(\vec{k}) + \mathbf{D}^e(\vec{k}). \quad (\text{A10})$$

In the present problem we are interested in calculating the change in the vibrational free energy of the crystal lattice due to the creation of a defect. This is done using the Green's function, which is most conveniently calculated in terms of the density of states of the phonon system. For the ideal crystal without a defect, a Brillouin zone integration is done to calculate the phonon density-of-states matrix Θ , from which the Green's function can be calculated via

$$\mathbf{G}^0(\omega^2) = \int_{\omega_{\min}^2}^{\omega_{\max}^2} \frac{\Theta(\omega'^2) - \Theta(\omega^2)}{\omega'^2 - \omega^2 - i\eta} d\omega'^2 + \Theta(\omega^2) \ln \left(\frac{\omega^2 - \omega_{\min}^2 - i\eta}{\omega^2 - \omega_{\max}^2 - i\eta} \right), \quad (\text{A11})$$

where the singularity in the integral has been explicitly removed. The change in the total density of states when a defect is introduced into the crystal can be deduced from Dyson's equation to obtain

$$\Delta\Theta(\omega^2) = \frac{1}{\pi} \text{Im} \frac{\partial}{\partial \omega^2} \ln \{ \det[1 - \mathbf{G}^0(\omega^2)\mathbf{V}] \}, \quad (\text{A12})$$

where \mathbf{V} is the perturbation potential. In the present case we use a strictly site-diagonal perturbation potential corresponding to the mass change due to the introduction of an isolated defect.

The partition function for the phonon system in the zinc blende lattice is given by

$$Z = \prod_i \left(\frac{\exp\left(\frac{-\hbar\omega_i}{2k_B T}\right)}{1 - \exp\left(\frac{-\hbar\omega_i}{k_B T}\right)} \right), \quad (\text{A13})$$

where ω_i are the normal modes of the system. The total vibrational entropy of the system is obtained from $S_{\text{total}}^{\text{vib}} = \frac{\partial}{\partial T} (k_B T \ln Z)$, which gives

$$S_{\text{total}}^{\text{vib}} = \sum_i \left\{ k_B \ln \left[\frac{1}{2} \sinh^{-1} \left(\frac{\hbar\omega_i}{2k_B T} \right) \right] + \frac{\hbar\omega_i}{2T} \coth \left(\frac{\hbar\omega_i}{2k_B T} \right) \right\}. \quad (\text{A14})$$

We convert the sum to an integral by the replacement

$$\sum_i \rightarrow \int_0^\infty \rho(\omega) d\omega = \int_0^\infty 2\rho(\omega^2) \omega d\omega \quad (\text{A15})$$

where $\rho(\omega^2) = \text{Tr} \Theta(\omega^2)$. We are interested in the change in the vibrational entropy upon formation of a defect, in which case we replace ρ by $\Delta\rho$ to obtain

$$S^{\text{vib}} = 2k_B \int_0^\infty \Delta\rho(\omega^2) \left\{ \frac{\hbar\omega}{2k_B T} \coth \left(\frac{\hbar\omega}{k_B T} \right) - \ln \left[2 \sinh \left(\frac{\hbar\omega}{2k_B T} \right) \right] \right\} \omega d\omega. \quad (\text{A16})$$

Similarly for the vibrational energy, we use the relationship $U^{\text{vib}} = -\frac{\partial}{\partial(1/k_B T)} \ln Z$ to obtain

$$U^{\text{vib}} = 2 \int_0^\infty \Delta\rho(\omega^2) \left[\frac{\hbar\omega}{2} \coth \left(\frac{\hbar\omega^2}{2k_B T} \right) \right] \omega d\omega, \quad (\text{A17})$$

with the change in the vibrational free energy $F^{\text{vib}} = U^{\text{vib}} - TS^{\text{vib}}$.

While the calculation of the vibrational terms in this paper was done using the general expressions above, it is interesting to examine the expression for the free energy in the high-temperature limit, which is appropriate for high growth and processing temperatures, and to examine the explicit temperature dependence of this term. In the high-temperature limit $\hbar\omega_{\max} \ll k_B T$ and F^{vib} reduces to

$$F^{\text{vib}} \approx 2k_B T \int_0^\infty \ln \left(\frac{\hbar\omega}{k_B T} \right) \rho(\omega) \omega d\omega. \quad (\text{A18})$$

This integral can be shown to be equal to

$$F^{\text{vib}} \approx k_B T n_m \ln \left(\frac{\hbar\bar{\omega}}{k_B T} \right), \quad (\text{A19})$$

where n_m is the number of phonon modes created or destroyed in the defect reaction of interest and $\bar{\omega}$ is an appropriately weighted frequency. What enters the calculation of the defect concentrations is $\exp(-F^{\text{vib}}/k_B T)$, which reduces to

$$\exp \left(\frac{F^{\text{vib}}}{k_B T} \right) \approx \left(\frac{k_B T}{\hbar\bar{\omega}} \right)^{n_m} = C T^{n_m} \quad (\text{A20})$$

in the high-temperature limit. For the defect reactions in Table I we obtain

$$n_m = \begin{cases} -3 & V_{\text{Hg}} \\ 3 & V_{\text{Te}} \\ 6 & \text{Hg}_{\text{Te}} \\ -6 & \text{Te}_{\text{Hg}} \\ +3 & \text{Hg}_I \\ -3 & \text{Te}_I \end{cases} \quad (\text{A21})$$

We will use this simple power-law dependence of Eq. (A20) to extract a simple power-law temperature dependence of the reaction constants.

- ¹ H. R. Vydyanath, *J. Electrochem. Soc.* **128**, 2609 (1981).
- ² D. A. Stevenson and M-F. S. Tang, *J. Vac. Sci. Technol. B* **9**, 1615 (1991), and references therein.
- ³ S. M. Johnson, D. R. Rhiger, J. P. Rosbeck, J. M. Peterson, S. M. Taylor, and M. E. Boyd, *J. Vac. Sci. Technol. B* **10**, 1499 (1992).
- ⁴ S. H. Shin, J. M. Arias, D. D. Edwall, M. Sandian, J. G. Pasko, and R. E. DeWames, *J. Vac. Sci. Technol. B* **10**, 1492 (1992).
- ⁵ R. S. List, *J. Electron. Mater.* **22**, 1017 (1993).
- ⁶ H. R. Vydyanath, *J. Appl. Phys.* **65**, 3080 (1989).
- ⁷ M. C. Chen, S. G. Parker, and D. F. Weirauch, *J. Appl. Phys.* **58**, 3150 (1985).
- ⁸ R. Sporken, M. D. Lange, S. Sivanathan, and J. P. Faurie, *Appl. Phys. Lett.* **59**, 81 (1991).
- ⁹ G. A. Baraff and M. Schlüter, *Phys. Rev. Lett.* **55**, 1327 (1985); *Phys. Rev. B* **30**, 1853 (1984); G. A. Baraff, E. O. Kane, and M. Schlüter, *ibid.* **21**, 5662 (1980).
- ¹⁰ D. B. Laks, C. G. Van de Walle, G. F. Neumark, and S. T. Pantelides, *Phys. Rev. Lett.* **66**, 648 (1991); D. B. Laks, C. G. Van de Walle, G. F. Neumark, P. E. Blöchl, and S. T. Pantelides, *Phys. Rev. B* **45**, 10965 (1992).
- ¹¹ J. Bernholc, N. O. Lipari, and S. T. Pantelides, *Phys. Rev. B* **21**, 3545 (1980); J. Bernholc and S. T. Pantelides, *ibid.* **18**, 1780 (1978); J. Bernholc, N. O. Lipari, and S. T. Pantelides, *Phys. Rev. Lett.* **41**, 895 (1978).
- ¹² S. B. Zhang and D. J. Chadi, *Phys. Rev. Lett.* **64**, 1789 (1990).
- ¹³ M. A. Berding, M. van Schilfgaarde, A. T. Paxton, and A. Sher, *J. Vac. Sci. Technol. A* **8**, 1103 (1990).
- ¹⁴ M. A. Berding, M. van Schilfgaarde, and A. Sher, *J. Vac. Sci. Technol. B* **10**, 1471 (1992).
- ¹⁵ M. A. Berding, A. Sher, and A.-B. Chen, *J. Appl. Phys.* **68**, 5064 (1990); *J. Vac. Sci. Technol. A* **5**, 3009 (1987).
- ¹⁶ J. T. Schick and C. G. Morgan-Pond, *J. Vac. Sci. Technol. A* **8**, 1108 (1990); C. G. Morgan-Pond and R. Raghavan, *Phys. Rev. B* **31**, 6616 (1985).
- ¹⁷ M. A. Berding, M. van Schilfgaarde, and A. Sher, *J. Electron. Mater.* **22**, 1005 (1993).
- ¹⁸ O. K. Andersen, O. Jepsen, and D. Glotzel, *Highlights of Condensed Matter Theory*, edited by F. Bassani *et al.* (North-Holland, Amsterdam, 1985), p. 59.
- ¹⁹ D. Langreth and D. Mehl, *Phys. Rev. B* **28**, 1809 (1983).
- ²⁰ M. van Schilfgaarde, A. T. Paxton, M. A. Berding, and M. Methfessel (unpublished).
- ²¹ F. A. Kröger, *The Chemistry of Imperfect Crystals* (J. Wiley & Sons, Inc., New York, 1964).
- ²² F. A. Kröger and H. J. Vink, in *Solid State Physics Vol. 3*, edited by F. Seitz and D. Turnbull (Academic Press, New York, 1956), p. 307.
- ²³ S. Krishnamurthy and A. Sher (unpublished).
- ²⁴ M. Methfessel and M. van Schilfgaarde (unpublished).
- ²⁵ U. von Barth and L. Hedin, *J. Phys. C* **5**, 1629 (1972).
- ²⁶ M. Methfessel, *Phys. Rev. B* **38**, 1537 (1988).
- ²⁷ R. Jones and A. Sayyesh, *J. Phys. C* **19**, L653 (1986).
- ²⁸ J. P. Hirth and J. Lothe, *Theory of Dislocations* (J. Wiley & Sons, Inc., New York, 1982), p. 49.
- ²⁹ E. R. Weber, H. Ennen, U. Kaufmann, J. Windscheif, J. Schneider, and T. Wosinski, *J. Appl. Phys.* **53**, 6140 (1982).
- ³⁰ C. G. Van de Walle, P. J. H. Denteneer, Y. Bar-Yam, and S. T. Pantelides, *Phys. Rev. B* **39**, 10791 (1989).
- ³¹ D. E. Cooper and W. A. Harrison, *J. Vac. Sci. Technol. A* **8**, 1112 (1990).
- ³² For a general discussion of the calculation of state degeneracy in semiconductors, see E. A. Guggenheim, *Proc. Phys. Soc. London Sect. A* **66**, 121 (1953), and references therein.
- ³³ G. L. Hansen and J. L. Schmit, *J. Appl. Phys.* **54**, 1639 (1983).
- ³⁴ G. L. Hansen, J. L. Schmit, and T. N. Casselman, *J. Appl. Phys.* **53**, 7099, (1982).
- ³⁵ There has been a recent attempt to predict the intrinsic carrier concentrations at temperatures up to 600 °C by M. Wienecke, M. Schenk, and H. Berger, *Semicond. Sci. Technol.* **8**, 299 (1993), although their analysis is suspect because of the temperature dependences they have assumed for the high-temperature mobility.
- ³⁶ P. N. Keating, *Phys. Rev.* **145**, 637 (1966).
- ³⁷ A. A. Maradudin, E. W. Montroll, G. H. Weiss, and I. P. Ipatova, in *Solid State Physics*, edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic Press, New York, 1971), Suppl. 3, p. 1.
- ³⁸ W. A. Harrison, *Electronic Structure and the Properties of Solids* (W. H. Freeman and Company, San Francisco, 1980).
- ³⁹ J. L. Meléndez and C. R. Helms (unpublished).
- ⁴⁰ H. F. Schaafe and J. H. Tregilgas, *J. Electron. Mater.* **12**, 931 (1983).

APPENDIX C

Hg_{0.8}Cd_{0.2}Te native defects: densities and dopant properties

M. A. Berding, M. van Schilfgaarde, and A. Sher

J. Electron. Mater. **22**, 1005 (1993)

Hg_{0.8}Cd_{0.2}Te Native Defects: Densities and Dopant Properties

M.A. BERDING, M. VAN SCHILFGAARDE, and A. SHER

SRI International, Menlo Park, CA 94025

We examine the native defect equilibrium in HgCdTe, including cation and anion vacancies, interstitials, and antisites in the analysis. A gradient correction to the local density functional has been added to the defect formation enthalpies calculated within the local density approximation, and preliminary predictions of the dominant ionization states are made. Temperature-dependent defect formation entropies and the temperature dependence of the pre-exponentials are incorporated into the calculation of the defect densities. Degenerate Fermi-Dirac statistics are used for the electronic equilibration, and the intrinsic reaction constant as a function of composition and temperature is calculated. We theoretically substantiate the doubly ionized mercury vacancy as the dominant defect in HgCdTe, and expect the doubly ionized mercury vacancy densities to be comparable in HgZnTe. We predict that tellurium antisites are donors and will be present for some annealing conditions in sufficient quantities to be measured and possibly to affect device performance.

Key words: Calculation of defect densities, HgCdTe, HgZnTe, native defects, Te antisites as donors

INTRODUCTION

The importance of native defects in HgCdTe is undisputed, with the dominant defect believed to be the double acceptor Hg vacancy.¹ The evidence for these defects is largely indirect and depends on their being ionized for observation. Neutral defects and compensating defects are more difficult to measure, and neither their densities nor even their presence is well established.

Our goal is to identify the important native defects in HgCdTe as a function of temperature and mercury pressure. Predictions of the absolute defect concentrations are difficult because of the accuracy required for reaction enthalpies and entropies that enter in exponentials. Our calculations although using a state-of-the-art method, are subject to a number of limitations, such as the supercell approximation that is used to calculate defect formation enthalpies. Despite these limitations, for which error bars can be esti-

mated, we find good agreement with experiment for the defect densities in narrow-gap HgCdTe. While there are some fitted parameters in the theory—for example, the temperature variation of the band gap—none have been chosen to fit the measured defect densities we are predicting. Thus, deviations must be ascribed to deficiencies in our approximations or to physical mechanisms that have not yet been incorporated.

In our previous work, we used the linearized muffin-tin orbital (LMTO) method within the atomic spheres approximation (ASA),² and later the full-potential (FP) Harris Foulkes approximation.³ The ASA substitutes a spheridized density for the true Hohenberg-Kohn density functional in the local density approximation. While the ASA is computationally fast, it cannot reliably predict atomic forces, and therefore lattice relaxations. Because of the elimination of the ASA shape approximation, in the FP calculations, we have predicted the breathing mode relaxations about the defect sites. Using the resulting defect formation enthalpies, in Ref. 3, we predicted

(Received October 12, 1992; revised January 13, 1993)

Table I. Formation Energies and Ionization States of the Native Point Defects

Defect	Formation Energy (eV)	Ionization State
V_{Hg}	1.93	shallow acceptor
V_{Te}	2.39	donor
Te_{Hg}	2.68	shallow donor
Hg_{Te}	0.75	deep acceptor
Te_i	4.47	shallow donor
Hg_i	1.75	shallow donor
ZnTe	4.66 (4.8)*	
CdTe	4.17 (4.4)*	
HgTe	3.37 (3.3)*	

Note: Formation energies refer to the neutral defect reactions in Eqs. 1–6 in text. See text for discussion of shallow and deep. *Experimental cohesive energies.

ratios of neutral native defects. In the present paper, we have made several important improvements to this previous work:

- A gradient correction to the local density functional⁴ has been included, which we find has a significant impact on the defect formation enthalpies;
- Absolute defect densities are predicted;
- The reaction constant for electron-hole pair production as a function of temperature and cadmium composition for degenerate Fermi-Dirac statistics is predicted; and
- The primary ionization states of the native defects have been tentatively identified and incorporated into the defect equilibrium.

The problem of predicting the defect concentrations is complicated by the fact that we are dealing with an alloy. The defect formation energies were calculated for HgTe, the primary component of the pseudobinary alloys HgCdTe and HgZnTe. Using a tight-binding model we have shown⁵ that the vacancy formation energy has a nonlinear dependence on the constituents in the near alloy environment, with the nonlinearity being most dramatic for the removal of the common-lattice atom (i.e. tellurium for HgCdTe and HgZnTe). The variation of the mercury vacancy formation energy, while less dramatic because the first-nearest neighbors are always tellurium, is as much as ~0.1 eV (depending on the constituents in the second-neighbor shell). We have not yet incorporated this level of detail into our first-principles calculations. Additionally, we have not included cadmium- or zinc-based defects (such as the cadmium antisite); because HgTe is the dominant constituent for compositions of technological importance, we believe this exclusion is justified. The alloy effects have been included in the present calculations via the composition dependence of the band gap which, because of its temperature dependence, has been extracted from experiment and the shape of the conduction and valence bands.

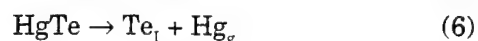
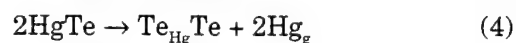
REACTION ENTHALPIES

We consider the following native defects: mercury vacancy V_{Hg} , tellurium vacancy V_{Te} , mercury antisite Hg_{Te} , tellurium antisite Te_{Hg} , mercury interstitial Hg_i , and tellurium interstitial Te_i . A couple of corrections have been added to the reaction enthalpies calculated within the Harris-Foulkes approximation to the FP-LMTO. First, the ASA is used to determine the ionization state of the defect—that is, whether it is a donor or acceptor. We have not yet determined the ionization energies of the defect, but we do indicate whether the state appears to be “deep” or “shallow.” We assume that only the shallow states are electrically active. Because we have not yet determined the ionization energies, we will for the present assume that they are zero for the shallow states—that is, that the donor and acceptor levels lie close to the conduction and valence band edges, respectively. We assume that the deep states are not electrically active. This approximation is likely to be good because, at high temperatures where the defect concentrations are equilibrated, carrier concentrations are high enough so that free carrier screening may effectively reduce the shallow-state activation energies. Details of the calculational method used to determine the ionization states of the various defects will be reported elsewhere.

The second correction to results involves the incorporation of a gradient correction to the local density functional. This correction has not yet been implemented in the FP-LMTO, and thus was done within the ASA. We expect that the FP gradient correction will be nearly equal to that from the ASA, because the density gradient is predominantly radial, the nonspherical components eliminated in the ASA being small. In Table I, we summarize the formation energies including the gradient correction for the neutral defect reactions in Eqs. 1–6 below. Also shown are the cohesive energies of the constituent compounds HgTe, CdTe, and ZnTe from a gradient-corrected self-consistent FP calculation; one can see that the agreement with experiments is quite good. The tentative identification of defect ionization states is also given in Table I.

DEFECT CONCENTRATIONS

Defect concentrations are determined using a quasi-chemical analysis of the defect formation reactions.⁶ We consider the following defect reactions



where we have chosen the HgTe unit cell and free

atomic mercury, Hg_g, as the reference states. Other reactions of interest can be obtained by taking linear combinations of these equations. For example, the neutral Schottky defect reaction is obtained by adding Eqs. 1 and 2 to obtain



or, as it is more commonly written,



The concentrations of the defects in Eqs. 1–6 can be obtained from the evaluation of the corresponding reaction constants:

$$K_{\text{V}_{\text{Hg}}} = [\text{V}_{\text{Hg}}] p_{\text{Hg}} \quad (9)$$

$$K_{\text{V}_{\text{Te}}} = [\text{V}_{\text{Te}}] / p_{\text{Hg}} \quad (10)$$

$$K_{\text{HgTe}} = [\text{HgTe}] / p_{\text{Hg}}^2 \quad (11)$$

$$K_{\text{TeHg}} = [\text{TeHg}] p_{\text{Hg}}^2 \quad (12)$$

$$K_{\text{Hg}_1} = [\text{Hg}_1] / p_{\text{Hg}} \quad (13)$$

$$K_{\text{Te}_1} = [\text{Te}_1] p_{\text{Hg}} \quad (14)$$

In these expressions, p_{Hg} is the mercury pressure in atmospheres and square brackets refer to concentrations per cubic centimeter. The evaluation of these reaction constants was discussed in our previous paper³ as well as in many standard texts.⁷

In the present calculations, we have used the formation energies from Table I and temperature-dependent entropies given in Ref. 3.

Ionized defect concentrations can be determined from the concentration of neutral defects from

$$[D^*] = (g_{D^*} / g_{D^0}) \exp((E_{D^0} - E_F) / k_B T) [D^0] \quad (15)$$

for donors and

$$[A'] = (g_{A'} / g_{A^0}) \exp((E_F - E_{A^0}) / k_B T) [A^0] \quad (16)$$

for acceptors. The dot and prime superscripts correspond to a positively and negatively charged species, respectively; g_X is the degeneracy of the state X; E_{D^0} and E_{A^0} are the positions of the first ionization levels for the donor and acceptor, respectively, in the one-electron picture; E_F is the Fermi energy; k_B is Boltzmann's constant; and T is the temperature in Kelvin. Similar expressions are obtained for the second ionization state.

In addition to the above equations for the determination of the native defect populations, we have the reaction for the generation of electron-hole pairs



The corresponding reaction constant is

$$K_{pn} = [h^*][e'] = pn \quad (18)$$

where $p = [h^*]$ and $n = [e']$, as in the usual notation.

In general, K_{pn} depends on the structure of the conduction and valence bands, the band gap energy, the Fermi energy, and the temperature variation of

these quantities. Several limits are often encountered in the evaluation of K_{pn} . First, when the conduction and valence bands are parabolic, although not necessarily isotropic, $E \sim k^2$, and the reaction constant can be written in terms of the Fermi-Dirac integrals as

$$K_{pn} = 4 \left(\frac{2\pi k_B T}{h^2} \right)^3 (m_h m_e)^{3/2} F_{1/2} \left(\frac{E_F - E_c}{k_B T} \right) F_{1/2} \left(\frac{E_v - E_F}{k_B T} \right) \quad (19)$$

where $F_{1/2}$ is the Fermi-Dirac function; E_c , E_v , and E_F are the conduction-band, valence-band, and Fermi energy, respectively; m_h and m_e are the hole and electron density-of-states effective masses, respectively; and h is Planck's constant. In the nondegenerate limit, this reduces to the familiar expression

$$K_{pn} = 4 \left(\frac{2\pi k_B T}{h^2} \right)^3 (m_h m_e)^{3/2} \exp \left(\frac{E_v - E_c}{k_B T} \right) \quad (20)$$

which is independent of the Fermi energy. For the general degenerate case, which will apply even at moderate temperatures for narrow-gap HgCdTe, K_{pn} will depend on E_F and therefore in general will not be independent of the presence of extrinsic carriers. Additionally, the assumption of parabolic bands may be poor for the narrow-gap materials, where the dispersion near the conduction band edge is rather more linear than parabolic,⁸ i.e. $E = \alpha k$. This case obtains with the Fermi-Dirac integral function of order 2 and

$$K_{pn} = 2 \left(\frac{2\pi k_B T m_h}{h^2} \right)^{3/2} \frac{2}{\pi^2} \left(\frac{k_B T}{\alpha} \right)^3 F_2 \left(\frac{E_F - E_c}{k_B T} \right) F_{1/2} \left(\frac{E_v - E_F}{k_B T} \right) \quad (21)$$

In the present program, we have used Eq. 21 to evaluate K_{pn} , with $m_h^* = 0.43$,

$$E_g(x, T) = E_c - E_v = -0.313 + 1.787x + 0.444x^2 - 1.237x^3 + 0.932x^4 + (0.667 - 1.714x + 0.760x^2)T / 1000 \quad (22)$$

taken from Ref. 9, and α chosen to yield good agreement with experimental values of the intrinsic carrier concentrations.¹⁰

The calculation of the intrinsic reaction constant for narrow-gap HgZnTe is more difficult to evaluate, given the more limited data base for evaluation of the high-temperature band gap, effective masses, and intrinsic carrier concentration.

From the zero-temperature band structures, the valence-band effective masses are found to be comparable for HgCdTe and HgZnTe with equal band gaps, while the conduction-band effective mass for HgZnTe is slightly larger than for HgCdTe.¹¹ Overall, though, we expect K_{pn} for the two materials to be comparable.

To calculate the native defect concentrations at a given temperature and mercury pressure, we have to determine the Fermi energy which satisfies the neutrality condition

$$2\sum [D_i^{\bullet\bullet}] + \sum [D_i^{\bullet}] + [h^{\bullet}] + 2\sum [A_i^{\bullet\bullet}] + \sum [A_i^{\bullet}] + [e^{\bullet}] \quad (23)$$

The concentrations of neutral defects which are independent of the Fermi energy can be solved for directly for a given p_{Hg} and T .

RESULTS

The low-temperature (77K) hole concentrations as a function of mercury pressure for various equilibration temperatures are shown in Fig. 1 for $x = 0.2$ $\text{Hg}_x\text{Cd}_{1-x}\text{Te}$. Shown for comparison are the experimental data from Vydyanath.¹ In the calculations, we allowed both atomic and electronic equilibrations at the high temperature at which the annealing takes place; we then assume the total defect concentrations are frozen-in upon quenching, and that at 77K only the electronic equilibrium is reestablished. In agreement with Vydyanath,¹ we conclude that the mercury vacancy is the dominant defect in HgCdTe ; and it behaves as a double acceptor, although we have not yet determined the activation energy. We find the tellurium antisite, which is a donor, is also an important defect.

In Fig. 2, we show a breakdown of the concentrations of various defects as a function of mercury pressure for the lowest and highest annealing temperatures of Fig. 1. The defect densities are those present at the annealing temperature, not at 77K, so as to show the atomic and electronic defects present during equilibration. Although the total number of each class of defects remains unchanged upon quenching, the relative concentrations of the neutral and ionized defects will be affected. From Fig. 2a, one can see that the material is intrinsic at the higher mercury pressures, and becomes extrinsic as the pressure is reduced and then highly compensated at the lowest mercury pressures. If the doubly ionized mercury

vacancy is the dominant defect, the intrinsic regime is characterized by $[h] \sim P_{\text{Hg}}^{-1}$, while the extrinsic, uncompensated, limit is characterized by $[h] \sim P_{\text{Hg}}^{-1/3}$. It is the conversion to extrinsic behavior and the compensating tellurium antisite donors that are responsible for the roll-off on the low-pressure side of the hole concentration curves in Fig. 1. For the material annealed at 655°C, shown in Fig. 2b, the material is nearly extrinsic (dominated by the doubly ionized mercury vacancy) for mercury pressures throughout the stability region, although compensation by the tellurium antisites becomes more important at the lower mercury pressures. The extrinsic behavior is evident in

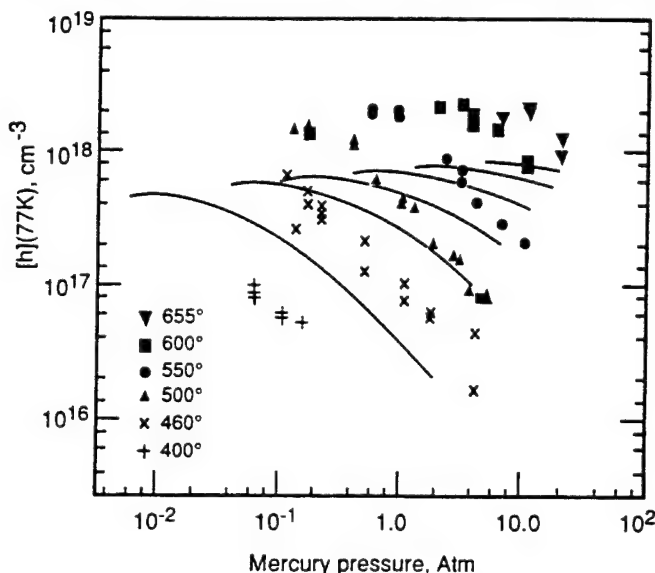


Fig. 1. Hole concentrations at 77K as a function of mercury pressure for material annealed at high temperatures. Experimental data were taken from Vydyanath.¹ Theory is shown as solid lines for pressures within the stability region at a given temperature.

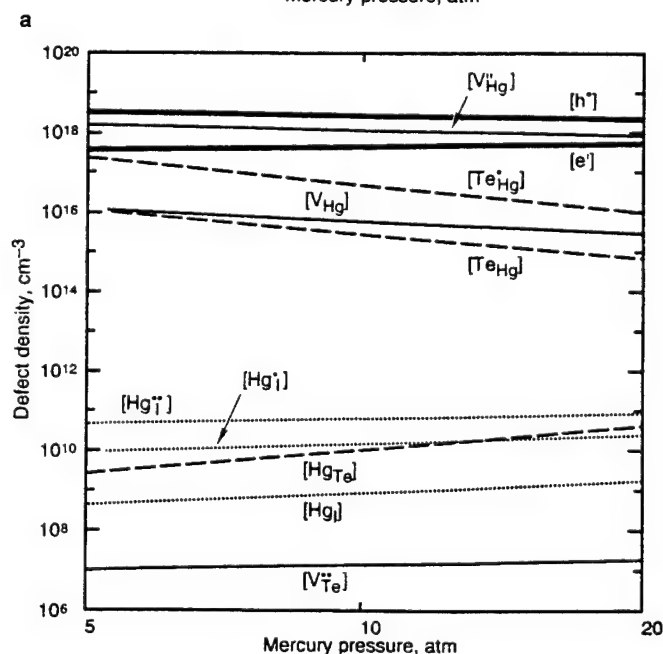
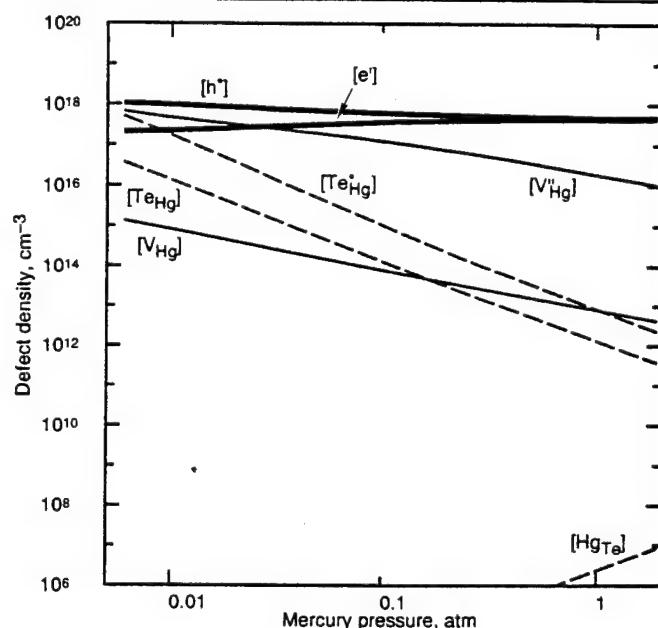


Fig. 2. Predicted defect densities present at the anneal temperature, as a function of mercury pressure for annealing temperatures of (a) 400°C and (b) 655°C.

Fig. 1 by the $[h] \sim P_{\text{Hg}}^{-1/3}$ behavior. As is evident in Fig. 1, the pressure at which the material goes from intrinsic to extrinsic shifts to higher mercury pressure as the annealing temperature is increased.

The discrepancies between our results and experiment can be attributed to a number of factors. First, we have calculated the formation energies for HgTe and applied them directly to HgCdTe, with the alloy taken into account in the calculation of the reaction constant for electron-hole pairs, and in the assumption that the defect ionization levels are at the band edges. Although this should be a reasonable approximation, we have shown that the vacancy formation energies are sensitive to the near-neighbor environment.⁵ The removal of this approximation will lead to a raising of the vacancy formation energy by as much as ~ 0.1 eV and, from examination of Fig. 1, will yield better agreement with experiment. Next, we have not taken into account the nonradial relaxation about the vacancy site and its effect on both the formation energy and entropy, nor have we yet included the effect of the entropy of local-mode softening of the ionized vacancy relative to the neutral vacancy. Additionally, the correct ionization energy must be incorporated into the calculation. Finally, we have used a band structure in calculating the reaction constant for electron-hole pairs which yields agreement with the intrinsic carrier concentrations and band gap at $T < 400\text{K}$; the band structure at higher temperatures is speculative, and we need further experiments or theory to confirm its validity.

At the annealing temperature, the material is intrinsic and all of the native defects are nearly completely ionized. While the total numbers of defects are frozen-in upon quenching from high temperature, the ratio of ionized to nonionized defects does change. At 77K, the material is extrinsic and, for material annealed at high temperatures ($T > 400^\circ\text{C}$), is dominated by the acceptor level of the mercury vacancy. Consequently, upon quenching, the ratio of ionized to nonionized defects decreases for acceptor defects and increases for donor defects.

Defect densities present after a low temperature ($\sim 250^\circ\text{C}$) annealing under mercury-saturated conditions are of technological interest. Our predictions for defect concentration as a function of $1/T$ for mercury pressures corresponding to mercury saturated conditions are shown in Fig. 3. At $T < 300^\circ\text{C}$, a mercury-saturated annealing step is effective in reducing the vacancy concentration to less than 10^{15} cm^{-3} . Because the density of tellurium antisites varies as P_{Hg}^{-2} , compared to P_{Hg} for the mercury vacancy, the antisites are most important on the tellurium-rich side of the stability region, and thus are less important for these mercury-saturated annealing.

While the mercury vacancy and tellurium antisite are the primary defects that will affect the electrical activity in Hg_{0.8}Cd_{0.2}Te, several other defects are of interest because of their role in diffusion, for example the mercury and tellurium interstitials. In Fig. 2, we see that the densities of mercury antisites and mer-

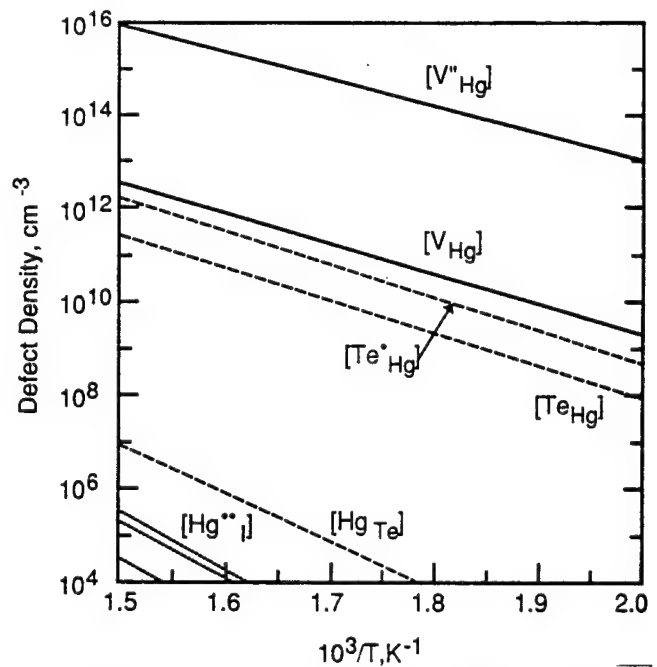


Fig. 3. Defect densities as a function of temperature for mercury-saturated annealing conditions.

cury interstitials are less than 10^{11} cm^{-3} , while the densities of tellurium vacancies and interstitials are completely negligible. The defect formation energies used for the calculation of the interstitial densities were from our ASA calculations without lattice relaxations, and therefore carry a larger uncertainty than the antisite and vacancy formation energies. We expect that improvement of the defect total-energy calculations will result in a potentially significant lowering of the formation energy, and consequently an increase in the interstitial defect concentrations. Even so, it is doubtful that we will find tellurium interstitials at densities large enough to account for the tellurium diffusion, which was believed to be via the tellurium interstitials.

While the tellurium antisites are donors and are predicted to be present in significant densities in HgCdTe, we do not predict that equilibrated tellurium antisites are responsible for the experimentally observed p-to-n conversion. Even if such a p-to-n conversion were to occur as a consequence of refinements in our calculations, the conversion would occur on the tellurium-rich side of the stability region. Because antisites are likely to be immobile, relatively high densities of tellurium antisites introduced at high temperatures during growth may persist through low-temperature annealings undertaken to remove mercury vacancies. In this case, the tellurium antisites could be responsible for the p-to-n conversion. A careful correlation of the n-type character with the high-temperature preparation conditions would help to confirm or negate this possibility.

Because of the similar magnitude for the electron and hole effective masses in narrow-gap HgCdTe and HgZnTe, we expect that the intrinsic reaction constants will be similar in the two materials. In addi-

tion, the mercury vacancy formation energy will be similar in the two materials for compositions low in cadmium or zinc, and therefore we expect the defect concentrations to be similar in the low-x materials with comparable band gaps. The tellurium antisite was found to be strained in the HgTe lattice^{2,3} and produced a compressive strain in the surrounding lattice. Because of the shorter bond lengths in the HgZnTe alloy, the strain associated with the tellurium antisite is larger than in HgCdTe with a comparable band gap; consequently, the formation energy is expected to be larger in HgZnTe. Thus, we expect the tellurium antisites to be less important in HgZnTe than in HgCdTe. If the tellurium antisite is responsible for the residual n-type character in low-temperature annealed material, lower n-type carrier concentrations will be realizable in HgZnTe.

CONCLUSIONS

We have calculated the low-temperature hole concentration as a function of annealing temperature and pressure for HgCdTe. We find good agreement with the fit to the experimental data. Remaining discrepancies can be attributed to relaxation and finite ionization energies in the vacancy formation energies, and uncertainties in K_{pn} at high temperatures. We conclude, in agreement with experiment, that the dominant defect is the doubly ionized mercury vacancy. We do not make any conclusions about the presence of the singly ionized Hg vacancy; we will examine the activation energies for both the singly and doubly ionized states in future work. We find that tellurium antisites are donors and are present in significant densities for low mercury pressures at temperatures above 350K. At lower temperatures

and high mercury pressures, at which annealing typically takes place, tellurium antisite equilibrium densities are negligible. However if their density remains at levels corresponding to high temperature because their low temperature diffusivities are low, then such antisites could affect mobilities and even be the residual donor. None of the other native defects are present in sufficient quantity to influence device characteristics.

ACKNOWLEDGMENT

This work has been supported by DARPA Contract MDA972-92-C-0053, NASA Contract NAS1-18226, and ONR contract N00014-89-K-132. Computational support provided by the Numerical Aerodynamical Simulation computing facility at NASA Ames Research Center.

REFERENCES

1. H.R. Vydyanath, *J. Electrochem. Soc.* 128, 2609 (1981).
2. M.A. Berding, M. van Schilfgaarde, A.T. Paxton and A. Sher, *J. Vac. Sci. Technol.* A 8, 1103 (1990).
3. M.A. Berding, M. van Schilfgaarde and A. Sher, *J. Vac. Sci. Technol.* B 10, 1471 (1992).
4. D. Lengreth and D. Mehl, *Phys. Rev. B* 28, 1809 (1983).
5. M.A. Berding, A. Sher and A.-B. Chen, *J. Appl. Phys.* 68, 5064 (1990); *J. Vac. Sci. Technol.* A, 5, 3009 (1987).
6. F.A. Kroger and H.J. Vink, *Solid State Physics Vol. 3*, eds. F. Seitz and D. Turnbull (Academic Press, New York 1956), p. 307.
7. For example, see F. Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, New York, 1965).
8. A.-B. Chen, M. van Schilfgaarde and A. Sher, this conference.
9. *Properties of Mercury Cadmium Telluride*, ed. J. Brice and P. Capper (INSPEC, New York, 1987), p. 105.
10. *ibid.*, p. 110.
11. M.A. Berding, S. Krishnamurthy, A. Sher and A.-B. Chen, *J. Vac. Sci. Technol.* A 5, 3014 (1987).

APPENDIX D

How dislocations affect transport

A. T. Paxton, A. Sher, M. Berding, M. van Schilfgaarde, M. W. Muller
J. Electron. Mater. **24**, 525 (1995)

How Dislocations Affect Transport

A.T. PAXTON

Department of Materials, University of Oxford, OX1 3PH, U.K.

A. SHER, M. BERDING, M. VAN SCHILFGAARDE

SRI International, Menlo Park, CA 94025

M.W. MULLER

Washington University, St. Louis, MO 63130

Dislocations crossing a junction in HgCdTe have little effect on detector responsivity, but are known to reduce the zero bias impedance R_0A and increase the leakage current, especially at low temperatures where R_0A is dominated by tunneling and generation/recombination processes. We have calculated the Coulomb and piezoelectric fields associated with dislocations in an attempt to interpret their effect on the junction's transport properties. Dislocation electric fields can affect transport since they are superimposed on the built-in and applied junction fields which control the currents. The screening of the fields in the neutral region is consistent with the dislocations' small effect on responsivity. Their impact in the space charge region is found to be significant and consistent with the nonlinear dependence of performance on dislocation density. The piezoelectric potential of the typical 60° dislocation in a sphalerite crystal, and the Coulomb potential of a dislocation crossing the junction plane other than normally, are angularly varying in the junction plane. Angular variation of the potentials can be qualitatively interpreted as an angular modulation of the potential barrier. Because of the nonlinear dependence of junction currents on the barrier (or the junction potential), the angular variation of the currents does not vanish upon averaging. We find that the range of the Coulomb potential is too small to account for a major portion of the experimentally reported performance degradation but may be responsible for the reduction of R_0A at cryogenic temperatures and low dislocation density, and that the longer range piezoelectric potential may be important. We also find that superposing the potentials of neighboring dislocations, because of the nonlinear dependence of junction leakage currents on junction potentials may account for the observed nonlinearity of performance degradation with dislocation density as measured by etch pit density.

Key words: Dislocations, HgCdTe, piezoelectric, transport properties

INTRODUCTION

The p-on-n $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ (MCT) double layer heterojunction (DLHJ) mesa-diode-based focal plane arrays (FPAs) shown in Fig. 1 are currently leading the art in long-wave infrared (LWIR) applications. The most effective structures to date are those grown

by liquid-phase epitaxy (LPE) on lattice-matched CdZnTe substrates. Typical values of active (n layer) and cap (p layer) compositions are $x = 0.224^1$ and $x \geq 0.28$, respectively. The n layers are typically indium-doped in the growth process, with carrier concentrations as low as permitted by series resistance and capacitance concerns, but usually realized at $> 2 \times 10^{15} \text{ cm}^{-3}$. The cap p⁺ layers are typically arsenic doped with $p \approx 2 \times 10^{17} \text{ cm}^{-3}$. The metallurgical p-n junction

(Received December 21, 1993; revised August 15, 1994)

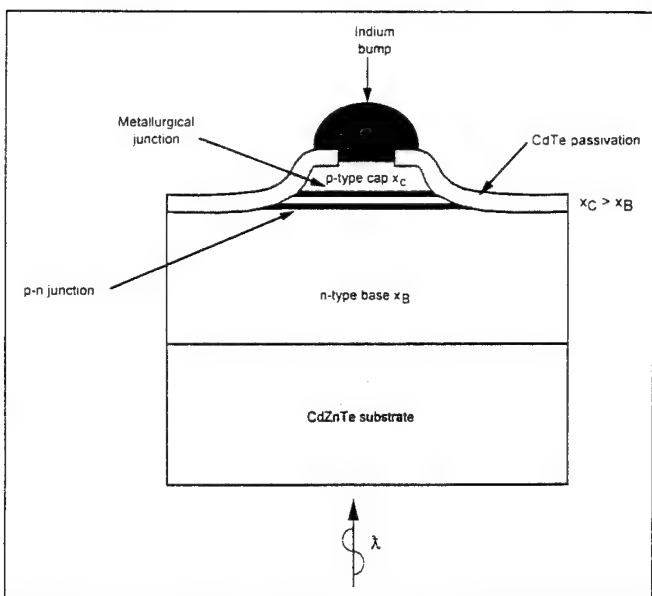


Fig. 1. Cross-sectional view of DLHJ mesa diode.

is positioned on the narrowgap side of the heterojunction, and the grading width is chosen to minimize the barrier in the valence band that occurs in such structures, as the barrier reduces the collection efficiency of photon-induced minority carriers (holes) in the n-layer from the back-lit array.

The virtues of the DLHJ are that the major absorption occurs in the n-type base layer where the photon-induced minority carriers are holes. The minority carrier lifetimes of holes in n-type materials are much longer than those of electrons in p-type material, which compensates for the mobility difference between electrons and holes. As a result, high quantum efficiencies are possible, and lower n-type carrier concentrations can be obtained. In addition, at least in principle, the generation-recombination (G-R) and tunneling dark currents are decreased in the heterojunction compared with those of homojunctions, and therefore the device noise is less. However, even if there is no tunneling current reduction, the wider bandgap still prevents interface-assisted recombinations from occurring at the metal contact with the p-cap layer.

There is ample experimental evidence²⁻⁵ that dislocations degrade the performance of MCT IR detectors; dislocations may be the most deleterious of the defects limiting the capabilities of the material. The measurements that lead to these conclusions relate various material parameters pertinent to detector performance to dislocation density. The important, interrelated parameters in the thin epitaxial layers suitable for the fabrication of LWIR photovoltaic devices are minority carrier lifetime, zero-bias junction impedance (R_0A product), and $1/f$ noise.

Dislocations are thought to influence the electronic properties of semiconductors primarily through their electric field. In this paper, we report the results of analyses of the space charge (Coulomb) and of the piezoelectric fields associated with dislocations in an

attempt to understand the mechanism through which they affect the responsivity and dark currents at low temperature.

SUMMARY OF EXPERIMENTAL RESULTS

In the work to be discussed, dislocation densities are measured by etch pit densities (EPDs). The etch pits on the film surface are associated with threading dislocations that are assumed to penetrate the films.

Shin et al.² have studied the relationship between dislocation density and minority carrier lifetime in epitaxial MCT films grown by LPE, metalorganic chemical vapor deposition (MOCVD), and molecular beam epitaxy (MBE). At small dislocation densities, minority carrier lifetime is not affected by the presence of dislocations, and the temperature and carrier concentration dependence of the lifetime suggests Auger recombination at high temperature (above 150K) and Shockley-Read recombination for $T < 150K$. However, even in samples with very low EPDs (down to pixels with no etch pits), the lifetime and the R_0A product exhibit great variability, sometimes exceeding an order of magnitude.

In these high-quality samples, the transition from Auger to trap-assisted recombination can occur at T as low as 60K³. We take this observation as evidence that the ultimate device performance limits are imposed by recombination centers not associated with threading dislocations. In the low temperature range of interest for IR detectors, for EPD densities above the mid- 10^5 range minority carrier lifetime decreases with dislocation density at an $EPD^{-2/3}$ to $EPD^{-3/4}$ rate. The observed behavior is consistent with dislocations acting as Shockley-Read recombination centers with a trap level 60 meV above the valence band edge and a trap density proportional to $EPD^{-3/4}$. Although not explicitly stated, the results reported suggest that dislocations active as recombination centers are likely to carry a core charge. Similar results have been reported by Baranskii et al.⁴ who find that mobile dislocations introduced in bulk MCT by indentation give rise to a trapping level at $E_v + 50$ meV and act as impurity getters. Johnson et al.⁵ have measured the dependence of R_0A in arrays of photovoltaic (PV) diodes on dislocation density by deliberately using plastic deformation to introduce dislocations in localized regions and carrying out the measurement on both high and low EPD diodes of the same array. They find that the resistance is relatively insensitive to EPD or decreases linearly at low densities, and quadratically at high densities. The nonlinear behavior sets in at small dislocation densities (mid- 10^5 cm^{-2} or less) at low temperature (40K), at higher densities ($>10^6$ cm^{-2}) at 78 and 120K.

Measurements of the temperature dependence of the leakage current of diodes with different dislocation densities show that the high temperature (diffusion regime) R_0A is not affected by the presence of dislocations, but that as the temperature is reduced below 150K, the R_0A values diverge, differing by several orders of magnitude at 40K and below. The

authors interpret the low temperature leakage current as G-R current. They adopt a model⁶ for estimating the reduction of the minority carrier diffusion length by recombination at dislocations, and then account for the inverse quadratic increase of the *dark* current at high dislocation densities by the *ad hoc* assumption that closely spaced dislocation pairs contribute a conductance that exceeds the sum of the conductances of two individual dislocations. We shall point out below that such nonlinear behavior may be a feature of tunneling currents associated with dislocations.

CHARGED DISLOCATIONS

The Coulomb fields associated with electrically charged dislocation cores can act as scattering and recombination centers and may attract or repel charged point defects. The presence and amount of charge on a dislocation core depends on the distribution of localized levels in the bandgap in the core and on the Fermi level in the semiconductor. Once the core structure of the dislocation has been determined, we can take a phenomenological view and consider only the electric field and potential generated by a given core charge. The effective radius of the field is taken to be that distance from the dislocation core within which the field is large enough to move a carrier the distance between dislocations in a minority-carrier lifetime as measured in dislocation-free material. In this way, we define a volume within which the lifetime and mobility are affected by the presence of a charged dislocation. If the affected volume surrounding all the dislocations is a significant fraction of the total volume, then their presence can affect lifetime and mobility and hence R_0A . This relates the depletion radius R or screening radius l to a threshold dislocation density n_D . The results of the calculation to follow, relating dislocation density, effective volume, and temperature, are summarized in Fig. 2.

The bandgap of the material sets the boundary condition for the potential at the core radius. We assume, conservatively, a core radius of 5 Å and a core edge potential of 0.1 eV. The core edge potential is limited to the bandgap by free carrier accumulation. It may of course be smaller, in which case the volume over which the dislocation's Coulomb potential can affect performance is reduced. We shall find that, except in the vicinity of the metallurgical junction as discussed below, the Coulomb potential is not a major factor affecting performance. Therefore, this conservative assumption is justified for providing a limit. We represent the dislocation core by a cylindrical charge density $-p$ in a cylinder of radius a in n -type material with donor density N_D .

We calculate the Coulomb field both using the depletion approximation and the exact space charge equation, which takes into account the thermal agitation of the carriers. The exact equation must be used if the screening radius is larger than the depletion radius. In the depletion approximation, the charge density is assumed to be equal to the donor density out

to a depletion radius R and zero outside. The potential, depletion radius, and core charge are determined by integrating the Poisson equation with the core edge potential as boundary condition. Using the values given above results in a plot of R as a function of carrier density shown in Fig. 2. It turns out that the radius is close enough to the Debye length to make the depletion approximation questionable at low temperatures, where the intrinsic carrier density is small.

The exact space charge equation is

$$\frac{1}{r} \frac{d}{dr} r \frac{d\phi}{dr} = \frac{2}{L^2} [\sinh(\phi - \phi_F) - \sinh \phi_F]$$

where

$$\phi = \frac{e\phi}{kT}, L^2 = \frac{\epsilon kT}{e^2 n_i}, \phi_F = \frac{E_F - E_i}{kT} = \sinh^{-1} \frac{N_D - N_A}{n_i}.$$

For ϕ small this is

$$\frac{1}{r} \frac{d}{dr} r \frac{d\phi}{dr} \approx \left(\frac{2}{L^2} \cosh \phi_F \right) \phi = \frac{2}{L_D^2} \phi$$

with the extrinsic Debye length

$$L_D^2 = \frac{L^2}{\cosh \phi_F} \approx \frac{\epsilon kT}{e^2 N_D} = \frac{2.77 \times 10^5}{N_D}, N_D \text{ in cm}^{-3}.$$

Then with $x = r/L_D$, we have $(1/x) (d/dx) x (d\phi/dx) = \phi$, the zero order modified Bessel equation. The solution of interest is the decreasing Hankel function $K_0(x)$.

The depletion approximation is only valid for $r < R - L_D$. Since R and L_D are comparable, the exact expression must be used. We let $\phi = CK_0(x)$ and evaluate C from $\phi(a)$ at $x = a/L_D$.

The electric field is

$$E = \frac{d\phi}{dr} = \frac{kT}{e} \frac{d\phi}{dr} = \frac{kT}{eL_D} \frac{d\phi}{dx} = \sqrt{\frac{nkT}{\epsilon}} CK_1(x)$$

where n is the carrier density.

During its lifetime τ a carrier moves a distance l , the

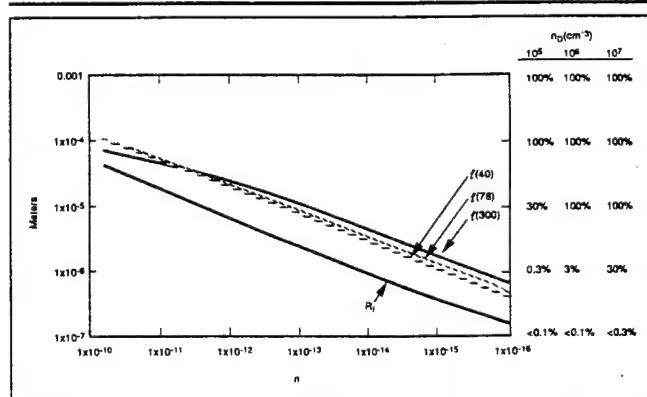


Fig. 2. Depletion radius R and screening radius l (meters) at 40, 70, and 300K as functions of carrier density in cm^{-3} . The table to the right of the figure converts the screening radius of the left-hand-axis to the percent volume of the material which is within the screening volume of a dislocation of various dislocation densities.

Table I. Mobility and Lifetime Values for Screening Radius l as a Function of n

T, K	μ , cm ² /Vs	τ , s
40	1.9×10^5	1×10^{-6}
78	9×10^4	3×10^{-7}
300	1×10^4	1×10^{-7}

mean free path in the Coulomb field, given by

$$\tau = \int_0^l \frac{dr}{v} = \int_0^l \frac{dr}{\mu E} = \int_0^l \frac{dr}{\mu \sqrt{\frac{nkT}{\epsilon}} CK_1 \left(\frac{r}{L_D} \right)}$$

where v is the carrier's velocity and μ its mobility. Given the lifetime and mobility, this expression can be solved for the free path l , that is to say, the distance the carrier drifts before recombining. We equate this distance with the effective range of the dislocation's field, which we identify with the screening radius. The screening radius should be distinguished from the smaller diffusion length an electron travels in a field free region in a lifetime. We show a plot of the screening radius l as a function of n for $T = 40, 77$, and 300 K in Fig. 2. To compute this plot, we have used the values of mobility and lifetime in Table I.

As expected, l turns out to be larger than the depletion radius, confirming the need for the more accurate screening theory. With the parameters we use, the calculation of the screening radius over five orders of magnitude of the carrier density, and over the temperature range from 40 to 300 K, results in values of l that range from about $5 L_D$ at $n = 10^{10} \text{ cm}^{-3}$ to $10 L_D$ at $n = 10^{15} \text{ cm}^{-3}$, resulting in a rate of decrease of l with carrier density at a rate slightly slower than $L_D^{-1/2}$. At the higher temperatures, the intrinsic carrier densities exceed N_D and the screening radius would be smaller.

The carrier density in the neutral region of the device material generally equals or exceeds 10^{14} cm^{-3} . The screening length at this carrier density, even at room temperature, is less than $6 \mu\text{m}$, and it is only about $3 \mu\text{m}$ at 40 K. Therefore, we can conclude that in the presence of mobile carriers, especially at cryogenic temperatures, the core charge fields have a limited range. This is also consistent with the observations of Johnson et al.⁵ that the responsivity of p on n heterojunction devices (dependent on diffusion in the neutral region) is little affected even by high dislocation densities.

The situation is less clear in the vicinity of the metallurgical junction of the p - n diode detector. In this region, the carrier density is reduced to the intrinsic value n_i under zero bias, and below that with reverse bias. We estimate a value of n_i about $5 \times 10^{11} \text{ cm}^{-3}$ for $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$ at 40 K, with a decrease of about a factor of two for every 5 mV of reverse bias. This reduction of the carrier density is confined to a small

region, and an assessment of its quantitative effect will require detailed computation. Nonetheless, it is worth observing that we estimate a screening radius of $45 \mu\text{m}$ for this carrier density. This is consistent with the experimental finding⁵ that $R_0 A$ at this temperature is already decreasing with dislocation densities in the 10^5 cm^{-2} range, corresponding to this order of separation between dislocation lines. The literature values of $n_i = 1.2 \times 10^{14} \text{ cm}^{-3}$ at 77 K and $n_i = 4.0 \times 10^{16} \text{ cm}^{-3}$ at 300 K⁷ correspond to progressively shorter screening lengths, which is also consistent with the experimental observation that the decrease in $R_0 A$ with dislocation density has a higher threshold at the higher temperature.

Because its form is geometrically simpler and more readily visualizable in this context, we have discussed screening entirely in connection with the effect of core charges. It should be kept in mind that similar, although quantitatively different, considerations apply to the screening of the piezoelectrically induced potentials, to be discussed next.

DISLOCATION STRAIN FIELDS AND PIEZOELECTRIC POTENTIALS

The MCT crystal structure lacks inversion symmetry and it, therefore, is piezoelectric. In a piezoelectric crystal, the strain field around a dislocation may generate an electric field. Conversely, an electric field applied to a piezoelectric crystal can produce a stress that may deform the crystal. This self-consistent field problem has been treated rigorously for a straight dislocation in a homogeneous insulating crystal (intrinsic wide gap semiconductor) by Saada.⁸ The general analysis was extended by Faivre and Saada⁹ to include the effects of screening by free carriers. Several qualitatively significant conclusions can be drawn from this work:

- The strains are only slightly affected by the electric field; that is, the inverse piezoelectric effect is small, and the electric field can be calculated directly without imposing the requirement of self-consistency.
- The electric field in the vicinity of the dislocation core can be quite large, and the dislocation can interact strongly with charges (point defects or mobile carriers) in its vicinity.
- In a semiconductor, the dislocation preferentially traps carriers of a particular sign.
- The estimated screening distance is of the order of several Debye lengths.

Booyens and Vermaak¹⁰ have applied Saada's analysis to $\langle 110 \rangle / \langle 111 \rangle$ dislocations in zinc-blende III-V semiconductors and Booyens and Basson¹¹ to CdTe and MCT. They do not explicitly include the effects of screening as analyzed by Faivre and Saada⁸ in these calculations. As a result, they find that an uncharged dislocation, in the absence of screening, generates a net radial field. They then suggest that the dislocation core is likely to acquire a charge that compensates for the radial field and so minimizes the total energy of the dislocation.

We have extended the analysis to the commonly observed 60° dislocations in intrinsic MCT. The dislocation lies along $[101]$ and has a Burgers vector \mathbf{b} at 60° to the dislocation line ξ . The edge component b of the Burgers vector is chosen as x_1 ; x_2 is the normal to the slip plane; and x_3 is ξ . The transformation matrix from principal axis coordinates is

$$T_{ij} = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 & -2 & 1 \\ \sqrt{2} & \sqrt{2} & \sqrt{2} \\ \sqrt{3} & 0 & \sqrt{3} \end{pmatrix}$$

The transformation of the piezoelectric tensor is

$$d'_{ijk} = T_{ip} T_{jq} T_{kr} d_{pqr}.$$

In cubic symmetry and principal axis coordinates, all the nonzero components of the piezoelectric tensor are equal: $d_{123} = d_{132} = d_{231} = d_{213} = d_{312} = d_{321} = d$. The nonzero components of the transformed tensor (d') are

$$\begin{aligned} d'_{133} &= d'_{313} = d'_{331} = -d'_{111} = \frac{2d}{\sqrt{6}}, \\ d'_{112} &= d'_{121} = d'_{211} = d'_{233} = d'_{323} = d'_{332} = \frac{d}{\sqrt{3}}, \\ d'_{222} &= \frac{2d}{\sqrt{3}}. \end{aligned} \quad (1)$$

The piezoelectrically induced bound charge density is¹²

$$\rho(r) = \frac{1}{\epsilon} d_{ijk} e_{jk,i} \quad \text{with} \quad e_{jk,i} \equiv \frac{\partial}{\partial x_i} e_{jk} \quad (2)$$

where ϵ is the isotropic dielectric constant, e_{jk} is the jk component of the strain tensor, and summation over repeated indexes is implied.

For a screw dislocation in linear anisotropic theory, the displacements are

$$u_1 = u_2 = 0, \quad u_3 = -\frac{b_3}{2\pi} \tan^{-1} A \left(\frac{x_1}{x_2} \right)$$

where b_3 is the screw component of the Burgers vector, and

$$A = \frac{\sqrt{c'_{44} c'_{55}}}{c'_{44}}; \quad c'_{44} = c_{44} - \frac{1}{3} H;$$

$$c'_{55} = c_{44} - \frac{1}{6} H; \quad H = 2c_{44} + c_{12} - c_{11}.$$

Here c_{ij} and c'_{ij} are the components of the elastic tensor in the principal axis and transformed coordinate systems, respectively. The nonzero components of the strain tensor are, using

$$e_{ij} = \frac{1}{2} (u_{i,j} + u_{j,i}); \quad e_{13} = e_{31} \quad \text{and} \quad e_{23} = e_{32};$$

all others zero.

Using Eq. (1) and Eq. (2) shows immediately that for this case $\rho = 0$.

Even in anisotropic linear elasticity (assuming the dielectric constant isotropic), the screw component of the 60° dislocation induces no piezoelectric effect. Therefore, pure screw $\langle 110 \rangle$ dislocations will also show no piezoelectric activity even if they are dissociated. We now proceed to consider the edge component. In order to arrive at a transparent expression, we will make the isotropic elastic approximation. This will not affect the order of magnitude calculations that follow.

For an edge dislocation, with ν Poisson's ratio

$$\begin{aligned} u_1 &= \frac{b}{2\pi} \left[\tan^{-1} \frac{x_2}{x_1} + \frac{x_1 x_2}{2(1-\nu)(x_1^2 + x_2^2)} \right], \\ u_2 &= -\frac{b}{2\pi} \left[\frac{1-2\nu}{4(1-\nu)} \ln(x_1^2 + x_2^2) + \frac{x_1^2 - x_2^2}{4(1-\nu)(x_1^2 + x_2^2)} \right], \\ u_3 &= 0. \end{aligned}$$

The nonzero strain components are

$$\begin{aligned} e_{11} &= \frac{b}{2\pi} \frac{(1-2\nu)x_2^3 + (3-2\nu)x_1^2 x_2}{2(\nu-1)(x_1^2 + x_2^2)^2}, \\ e_{22} &= \frac{b}{2\pi} \frac{(1-2\nu)x_2^3 - (1+2\nu)x_1^2 x_2}{2(\nu-1)(x_1^2 + x_2^2)^2}, \\ e_{12} &= \frac{b}{2\pi} \frac{x_1 x_2^2 - x_1^3}{2(\nu-1)(x_1^2 + x_2^2)^2}. \end{aligned}$$

From Eq. (2) and Eq. (1),

$$\begin{aligned} \rho(r) &= \frac{1}{\epsilon} (d'_{111} e_{11,1} + 2d'_{112} e_{12,1} \\ &\quad + d'_{211} e_{11,2} + d'_{222} e_{22,2}) \end{aligned}$$

so, substituting the components of the piezoelectric tensor from Eq. (1) and evaluating the strain tensor gradient, we find

$$\begin{aligned} \rho(x_1, x_2) &= \frac{d}{\sqrt{3}\epsilon} (-\sqrt{2}e_{11,1} - 2e_{12,1} - e_{11,2} + 2e_{22,2}) \\ &= \frac{bd}{2\pi\sqrt{3}\epsilon} \left[\frac{(2\nu-3)x_2^4 - 2\sqrt{2}(2\nu+1)x_1 x_2^3}{2(\nu-1)(x_1^2 + x_2^2)^3} \right. \\ &\quad \left. + \frac{30x_1^2 x_2^2 + 2\sqrt{2}(3-2\nu)x_1^3 x_2 - (2\nu+7)x_1^4}{2(\nu-1)(x_1^2 + x_2^2)^3} \right]. \end{aligned}$$

When this expression is converted to polar coordinates in the $x_1 x_2$ plane (xy plane), it becomes a somewhat simpler expression with angular terms having quadrupole and hexadecapole form:

$$\rho = \frac{bd}{2\pi\sqrt{3}\epsilon}(\rho_1 + \rho_2 + \rho_3 + \rho_4 + \rho_5)$$

with

$$\rho_1 = \frac{(2\nu-3)\cos 4\theta - 4\cos 2\theta + 3}{2(\nu-1)8r^2},$$

$$\rho_2 = -\frac{2\sqrt{2}(2\nu+1)\sin 4\theta - 2\sin 2\theta}{2(\nu-1)8r^2},$$

$$\rho_3 = \frac{30}{2(\nu-1)}\frac{1-\cos 4\theta}{8r^2},$$

$$\rho_4 = \frac{2\sqrt{2}(3-2\nu)\sin 4\theta + 2\sin 2\theta}{2(\nu-1)8r^2},$$

$$\rho_5 = -\frac{(2\nu+7)\cos 4\theta + 4\cos 2\theta + 3}{2(\nu-1)8r^2}$$

so

$$\rho = \frac{bd}{4\pi\sqrt{3}\epsilon(1-\nu)}\frac{1}{r^2}.$$

$$[5\cos 4\theta - \sqrt{2}\sin 4\theta + 2(1+\nu)\cos 2\theta - \sqrt{2}(1-2\nu)\sin 2\theta]$$

$$\equiv \frac{C}{r^2} \sum_{n=2,4} [a_n \cos(n\theta) + b_n \sin(n\theta)]$$

which defines a useful constant C , and four parameters depending only on Poisson's ratio.

This simple expression reveals the charge density to have an inverse square decay and quadrupole and hexadecapole terms only. Note in particular that there is no monopole term and hence no net piezoelectric charge induced on the dislocation line. This rather transparent expression has not been revealed in previous work (Refs. 10, 11).

To obtain the piezoelectric potential, we need to solve the Poisson equation $\nabla^2\Phi = -4\pi\rho/\epsilon$ with this charge distribution as the source. In two dimensions, the formal solution of the Poisson equation is

$$\Phi(r, \theta) =$$

$$\frac{1}{\pi} \int_0^\infty r' dr' \int_0^{2\pi} d\theta' \left[\ln \frac{1}{r_>} + \sum_{m=1}^\infty \frac{1}{m} \left(\frac{r_<}{r_>} \right)^m \cos m(\theta - \theta') \right] \rho(r')$$

$$= \frac{C}{\pi} \sum_{m=1}^\infty \frac{1}{m} \int_0^{2\pi} d\theta' [\sin(m\theta)\sin(m\theta') + \cos(m\theta)\cos(m\theta')]$$

$$\times \sum_{n=2,4} [a_n \cos(n\theta') + b_n \sin(n\theta')] \int_0^\infty \frac{r' dr'}{r'^2} \left(\frac{r_<}{r_>} \right)^m$$

where $r_>$ and $r_<$ are the greater and lesser, respectively, of r and r' . In analogy with the solution for the elastic field, the integral diverges at the origin because of the r^{-2} dependence of ρ . To deal with this difficulty, we adopt the same procedure that is used to avoid the

divergence of the elastic energy of a dislocation at the origin in continuum theory. We set cutoffs at an inner radius, R_1 , and an outer radius, R_2 . The cutoff at R_1 removes the core singularity and the cutoff at R_2 allows ρ to be expanded in multipoles. Later, we will identify R_1 as the core radius and we will let $R_2 \rightarrow \infty$. We now have

$$\Phi(r, \theta) = \frac{C}{\pi} \sum_{n=2,4} \frac{\pi}{n}.$$

$$\left[a_n \cos(n\theta) + b_n \sin(n\theta) \left(\frac{1}{r^n} \int_{R_1}^r (r')^{n-1} dr' + r^n \int_r^{R_2} (r')^{-n-1} dr' \right) \right]$$

$$= C \left\{ \frac{1}{4} [a_4 \cos(4\theta) + b_4 \sin(4\theta)] \left[\frac{1}{2} - \frac{1}{4} \left(\frac{R_1}{r} \right)^4 - \frac{1}{4} \left(\frac{r}{R_2} \right)^4 \right] \right. \\ \left. + \frac{1}{2} [a_2 \cos(2\theta) + b_2 \sin(2\theta)] \left[1 - \frac{1}{2} \left(\frac{R_1}{r} \right)^2 - \frac{1}{2} \left(\frac{r}{R_2} \right)^2 \right] \right\}$$

for $r < R_2$. For $r > R_2$, we have

$$\Phi(r, \theta) = C \left\{ \frac{1}{4} [a_4 \cos(4\theta) + b_4 \sin(4\theta)] \frac{1}{4} \left[\left(\frac{R_2}{r} \right)^4 - \left(\frac{R_1}{r} \right)^4 \right] \right. \\ \left. + \frac{1}{2} [a_2 \cos(2\theta) + b_2 \sin(2\theta)] \frac{1}{2} \left[\left(\frac{R_2}{r} \right)^2 - \left(\frac{R_1}{r} \right)^2 \right] \right\}$$

which is a multipole potential. The two solutions are matched at R_2 . We can now let $R_2 \rightarrow \infty$ and neglect the terms in R_1 , which amount to a small correction near the core, to get a potential valid for $r \gg R_1$

$$\Phi(r, \theta) =$$

$$C \left\{ \frac{1}{8} [a_4 \cos(4\theta) + b_4 \sin(4\theta)] + \frac{1}{2} [a_2 \cos(2\theta) + b_2 \sin(2\theta)] \right\}$$

which is independent of r .

This rather surprising behavior is a consequence of the inverse square decay of the charge density. In this connection, we can pursue the analogy with the well-known elastic solution for an edge dislocation by pointing out that while the elastic energy per unit length is given by

$$\frac{\mu b^2}{4\pi(1-\nu)} \left(\ln \frac{R_2}{R_1} - \frac{1}{2} \right)$$

(where μ is the isotropic shear modulus), the piezoelectric energy per unit length is (taking $\nu = 0.3$)

$$\frac{9\pi^2 C^2}{1.6} \left(\ln \frac{R_2}{R_1} - \frac{1}{8} \right)$$

which is six orders of magnitude smaller, justifying our neglecting the reverse piezoelectric effect. In both cases, the first term comes from the energy in an

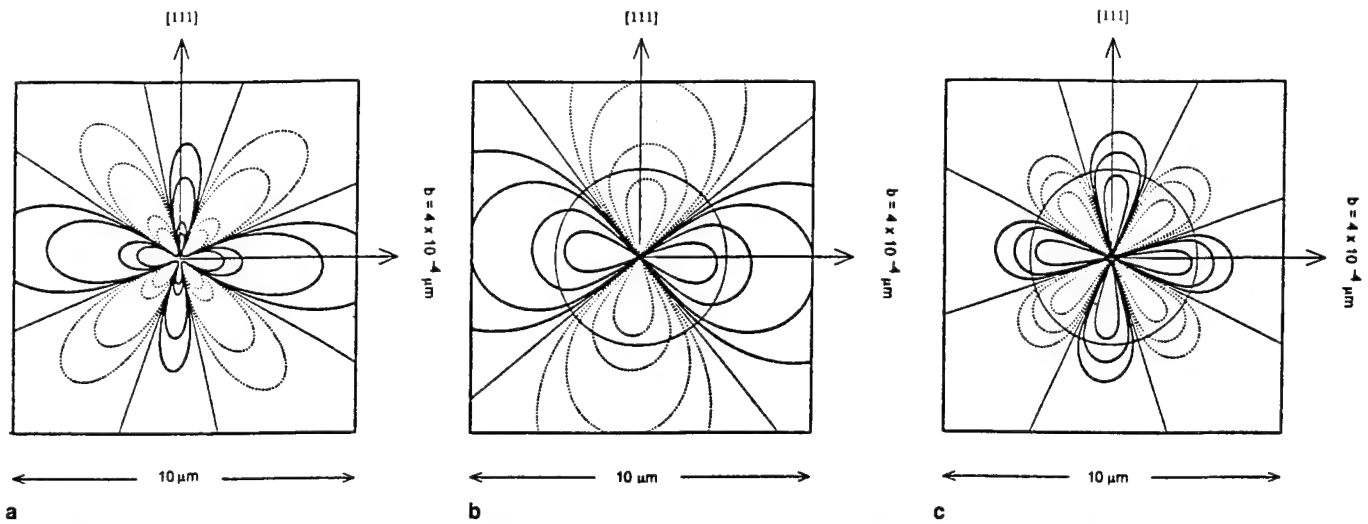


Fig. 3. Piezoelectric properties of a 60° dislocation in $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$: (a) Piezoelectric charge normal to the dislocation line. Solid contours are $1, 0.5, 0.1$, and $0.05 \times 10^{12} \text{ e/cm}^3$ going out radially; dotted contours are the negative of these; (b) quadrupole component of the piezoelectric potential, and (c) hexadecapole component of the piezoelectric potential.

infinite cylinder integrated to R_2 which diverges logarithmically, and the second term is the effect of the cylindrical surface bounding the dislocation in a finite crystal (the image term).

For a good numerical calculation, we would need the piezoelectric constant of the alloy, which has not been determined. For an order-of-magnitude estimate, we can use the piezoelectric constant of CdTe , $3 \times 10^{-6} \text{ coul cm}^{-2}$. The Burgers vector is $4 \times 10^{-8} \text{ cm}$, and Poisson's ratio can be taken as 0.3. Using these values, we have computed the charge, potential, and field plots of Figs. 3a–3c.

It should be understood that because of the symmetry of the strain field associated with a straight dislocation, several of the possible sources of the piezoelectric potential are identically zero. In the reduced symmetry of curved dislocations, kinks, and jogs, these components of the potential do not vanish and may be significant.

AVERAGING NONLINEAR EFFECTS

The results we have obtained afford a qualitative insight into the effect of the piezoelectric potential on the characteristics of a pn junction threaded by a dislocation. The angular variation of the potential will alternately add to and subtract from the built-in potential of the junction, so that the barrier seen by a carrier crossing the junction will depend on its trajectory. The modification of the built-in potential due to the angularly varying piezoelectric potential can increase or lower the barrier seen by a carrier crossing the junction in a direction other than perpendicularly. Because of the strongly nonlinear dependence of both diffusion and tunneling currents on the barrier, the effect of such a barrier variation does not average to zero and the currents can be enhanced by a large factor. To illustrate the principle, we choose as an example the average tunneling probability for carriers crossing a narrow semiconductor junction.

The band-to-band tunneling probability has the

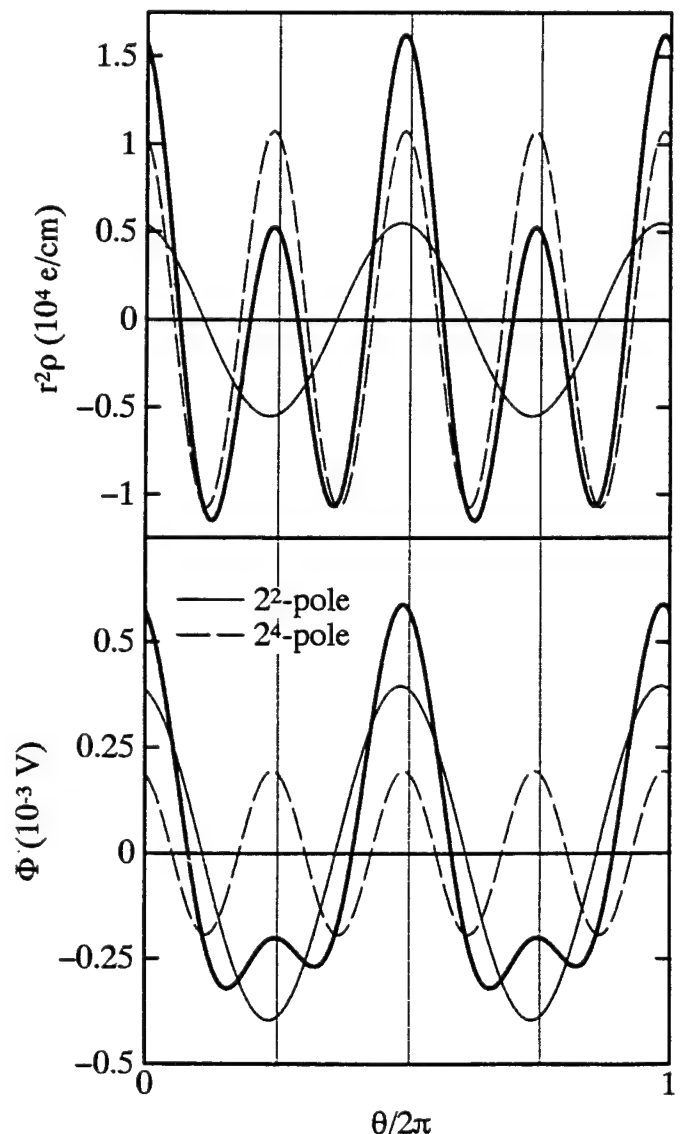


Fig. 4. Angular variation of the piezoelectric charge and potential normal to a 60° dislocation line.

form $p = \exp(-CB^{3/2})$ where B is the barrier and where C is a constant containing all the other material parameters.

We crudely represent the effect of the dislocation potential, which makes the tunneling barrier a function of the carrier's trajectory, as an angular variation of the barrier of the form $B = B_0(1 + \Delta \sin \theta)$ (it will be apparent presently that the detailed form of the angular variation is irrelevant for this argument). It is likely that this representation of the piezoelectric potential introduced by a dislocation can be justified by a WKB calculation.

If Δ is small compared with unity, we have $B^{3/2} = B_0^{3/2}[1 + (3/2)\Delta \sin \theta]$ (this approximation is taken purely to simplify the arithmetic and has no effect on the argument). We can now compute an average of p

$$\begin{aligned} \langle p \rangle &= \frac{1}{2\pi} \int_0^{2\pi} \exp\left[-CB_0^{3/2}\left(1 + \frac{3}{2}\Delta \sin \theta\right)\right] d\theta = \\ &= \frac{p_0}{2\pi} \int_0^{2\pi} \exp\left[-\frac{3}{2}CB_0\Delta \sin \theta\right] d\theta = p_0 I_0\left(\frac{3}{2}CB_0\Delta\right) \end{aligned}$$

where $I_0(x)$ is the modified zero order Bessel function and p_0 is the tunneling probability in the absence of any dislocations. From this last form, it is apparent that the average tunneling probability could have been written as

$$\langle p \rangle = p_0 I_0\left(\frac{3}{2}\Delta \ln|p_0|\right).$$

It is apparent that although the fractional barrier "angular modulation" Δ is small, the argument of the Bessel function need not be, and the factor multiplying the unmodulated barrier transition probability can be quite large. It is also evident that any other angular variation would lead to very similar results. The angular variation Δ can be estimated from the angular potential plots of Fig. 4 to be roughly 0.05–0.1. The parameters determining p_0 (effective mass, doping, temperature, etc.) cover a wide range. For typical values of these parameters, the multiplying

Bessel function ranges from near unity to about ten.

As another consequence of the type of nonlinearity examined here, the effect of overlapping piezoelectric potentials of neighboring dislocations is expected to be enhanced over what might be expected from linear superposition. This may be a possible clue to the observed inverse faster than linear falloff of R_0A with dislocation density.

CONCLUSIONS

We have evaluated the electric fields associated with dislocation charges and strain fields and estimated their effects on junction performance. We conclude that the fields associated with core charges alone are not likely to account for the observed degradation of junction performance. The longer range piezoelectric potentials, which we have calculated explicitly for the common 60° dislocation, are likely to have a larger effect, especially on R_0A through the tunneling leakage current. The effect of the nonlinearity of the junction characteristic on R_0A is examined and is proposed as a mechanism to account for observed nonlinear behavior.

ACKNOWLEDGMENT

Work supported by DARPA contract MDA972-92-C-5003.

REFERENCES

1. G.L. Hansen, J.L. Schmit and T.N. Casselman, *J. Appl. Phys.* 53, 7099 (1982).
2. S.H. Shin, J.M. Arias, D.D. Edwall, M. Zandian, J.G. Pasko and R.E. DeWames, *J. Vac. Sci. Technol. B* 10, 1492 (1992).
3. R.E. DeWames, private communication 1993.
4. P.I. Baranskii, A.E. Belyaev, O.P. Gorodnicheii and S.M. Komirenko, *Sov. Phys. Semicond.* 24, 73 (1990).
5. S.M. Johnson, R. Righer, J.P. Rosbeck, J.M. Peterson, S.M. Tailor and M.E. Boyd, *J. Vac. Sci. Technol. B* 10, 1499 (1992).
6. J.C. Zolper and A.M. Barnett, *IEEE Trans. Elect. Dev.* 37, 478 (1990).
7. J. Brice and P. Capper, eds., *The Properties of Mercury Cadmium Telluride, EMIS Data Reviews, Series No. 3*, (New York: INSPEC, 1987).
8. G. Saada, *Phys. Stat. Sol. (B)* 44, 717 (1971).
9. G. Faivre and G. Saada, *Phys. Stat. Solidi. (B)* 52, 127 (1972).
10. H. Booyens and J.S. Vermaak, *J. Appl. Phys.* 50, 4302 (1979).
11. H. Booyens and J.H. Basson, *Phys. Stat. Solidi (A)* 85, 243 (1984).
12. P.A. Fedders, *J. Appl. Phys.* 54, 1804 (1983).

APPENDIX E

Dislocations and junction performance in MCT

M. W. Muller A. T. Paxton, A. Sher, M. Berding, M. van Schilfgaarde

Proceedings of the IRIS Materials Meeting, August 1993

DISLOCATIONS AND JUNCTION PERFORMANCE IN MCT*

August 1993

M.W. Muller[†], A.T. Paxton[§], A. Sher, M. Berding, M. van Schilfgaarde
SRI International, Menlo Park, CA 94025

ABSTRACT

Dislocations crossing a junction in HgCdTe have little effect on detector responsivity, but are known to reduce the zero bias impedance R_0A and increase the leakage current, especially at low temperatures where R_0A is dominated by tunneling and GR processes^{1,2}. We have calculated the core charge and piezoelectric-induced fields associated with 60° dislocations since they are superimposed on the built-in and applied junction fields that control the currents. The screening of the dislocation fields in the neutral region is nearly complete and is consistent with the dislocations' small effect on responsivity. Their impact in the space charge region is found to be significant and consistent with the nonlinear dependence of performance on dislocation density, as characterized by R_0A .

1. INTRODUCTION

The p-on-n $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ (MCT) double-layer heterojunction (DLHJ) focal plane arrays (FPAs) shown in Figure 1 are currently leading the art in long-wave infrared (LWIR) applications. The most effective structures to date are those grown by liquid-phase epitaxy (LPE) on lattice-matched CdZnTe substrates. Typical values of active (n layer) and cap (p layer) compositions are $x = 0.224$ and $x \geq 0.28$, respectively. The n layers are typically indium-doped

* This work was supported by ARPA Contract MDA972-92-C-0053.

† Permanent address: Washington University, St. Louis, MO 63130

§ Present Address: Department of Materials, University of Oxford, OXI 3PH, UK

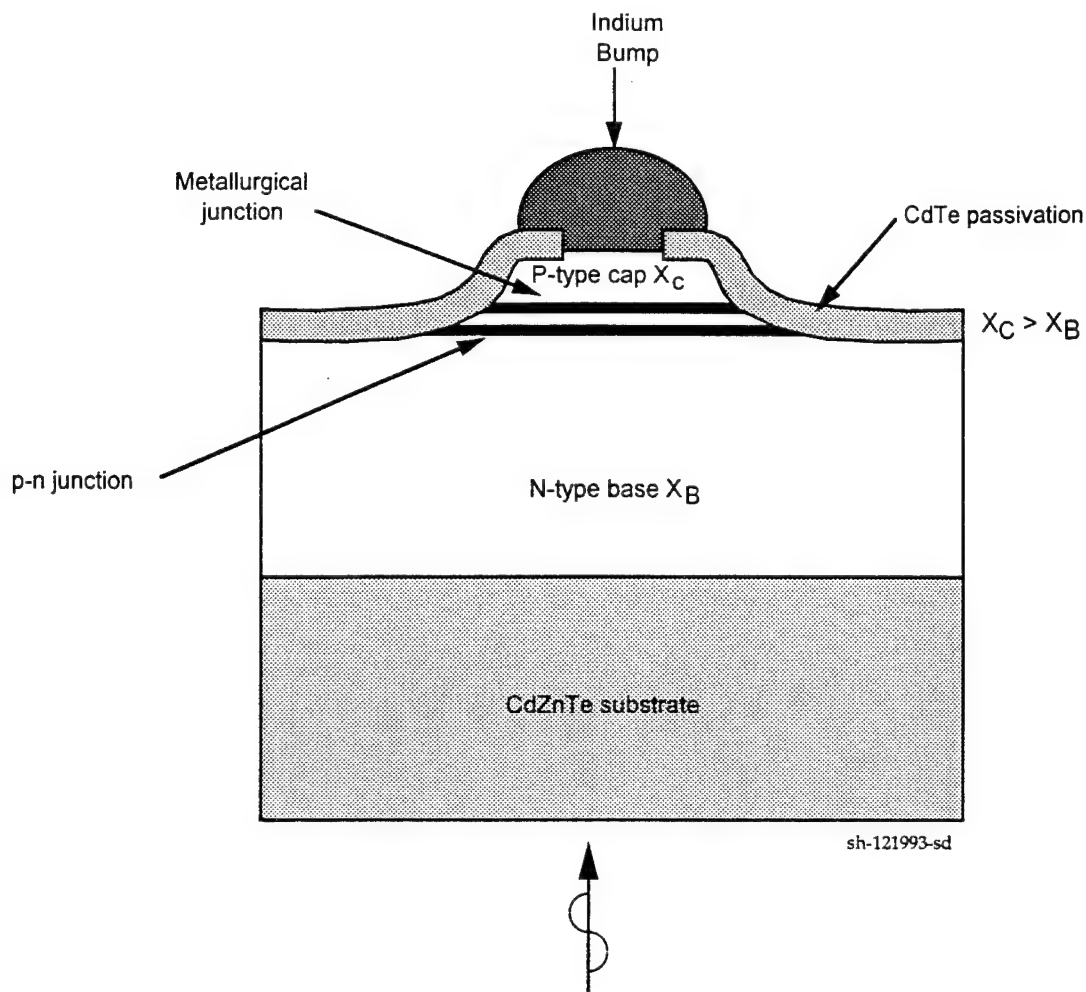


Figure 1. Cross-sectional view of DLHJ mesa diode

Source: R.E. DeWames et al., *An Assessment of HgCdTe and GaAs/GaAlAs Technologies for LWIR Infrared Imagers*, SPIE, Vol. 1735, p. 6, 1992.

in the growth process, with carrier concentrations as low as permitted by series resistance and capacitance concerns, but usually realized at $> 2 \times 10^{15} \text{ cm}^{-3}$. The cap p^+ layers are typically arsenic-doped with $p \cong 2 \times 10^{17} \text{ cm}^{-3}$. The metallurgical p-n junction is positioned on the narrow-gap side of the heterojunction, and the grading width is chosen to minimize the barrier in the valence band that occurs in such structures³, as the barrier reduces the collection efficiency of photon-induced minority carriers (holes) in the n layer from the back-lit array.

There is ample experimental evidence that dislocations degrade the performance of DLHJ detectors and may be the defects currently limiting the capabilities of the devices. The measurements that relate various material parameters pertinent to detector performance to dislocation density have been reported by several groups.^{1,2} The important, interrelated

parameters are minority carrier lifetime, zero-bias junction impedance (R_0A product), and $1/f$ noise.

Dislocations are thought to influence the electronic properties of semiconductors primarily through their electric fields. In this paper we report the results of analyses of the space charge and piezoelectric fields associated with dislocations in an attempt to understand the mechanism through which they affect the responsivity and dark currents at low temperature.

2. SUMMARY OF EXPERIMENTAL RESULTS

In the work to be discussed, dislocation densities are measured by etch pit densities (EPDs). The etch pits on the film surface are associated with threading dislocations that are assumed to penetrate the films. The relation of these dislocations to misfit dislocations and loops formed at growth interfaces is not fully understood in all cases.

Shin *et al.*² have studied the relationship between dislocation density and minority carrier lifetime in epitaxial MCT films grown by LPE, metalorganic chemical vapor deposition (MOCVD), and molecular beam epitaxy (MBE). At small dislocation densities (as measured by etch pit density) the minority carrier lifetime is not affected by the presence of dislocations, and the temperature and carrier concentration dependence of the lifetime suggests Auger recombination at high temperature (above 150 K) and Shockley-Read recombination for $T < 150$ K. However, even in samples with very low EPDs (down to pixels with no etch pits), the lifetime and the R_0A product exhibit great variability, sometimes exceeding an order of magnitude. In these high-quality samples, the transition from Auger to trap-assisted recombination can occur at temperatures as low as 60 K.⁴ We take this observation as evidence that recombination centers not associated with threading dislocations may be the defects imposing the ultimate device performance limits. In the low temperature range of interest for IR detectors, for EPD densities above the mid 10^{-5} range, minority carrier lifetime decreases with dislocation density at an $EPD^{-2/3}$ to $EPD^{-3/4}$ rate. Shin *et al.*² note that the observed behavior is consistent with dislocations acting as Shockley-Read recombination centers with a trap level 60 meV above the valence band edge and a trap density proportional to $EPD^{-3/4}$. Although not explicitly stated in this paper, the results reported suggest that dislocations active as recombination centers are likely to carry a core charge. Similar results have been reported by Baranskii *et al.*⁵ who find that mobile dislocations introduced in bulk MCT by indentation give rise to a trapping level at $E_v + 50$ meV.

Johnson *et al.*¹ have measured the dependence of R_0A in arrays of photovoltaic diodes on dislocation density by deliberately using plastic deformation to introduce dislocations in localized regions and carrying out the measurement on both high and low EPD diodes of the same array. They find that the resistance is relatively insensitive to EPD at the lowest dislocation densities, decreases linearly at higher densities, and decreases quadratically at the highest densities. The nonlinear behavior sets in at small dislocation densities (mid 10^5 cm^{-2} or less) at low temperature (40 K) and at higher densities ($>10^6 \text{ cm}^{-2}$) at 78 and 120 K. Measurements of the temperature dependence of the leakage current of diodes with different dislocation densities show that the high temperature (diffusion regime) R_0A is not affected by the presence of dislocations, but that as the temperature is reduced below 150 K the R_0A values diverge, differing by several orders of magnitude at 40 K and below. The authors interpret the low temperature leakage current as generation-recombination current. They adopt a model (Zolper and Barnett⁶) for estimating the reduction of the minority carrier diffusion length by recombination at dislocations, and then account for the inverse quadratic increase of the short circuit current at high dislocation densities by the *ad hoc* assumption that closely spaced dislocation pairs contribute a conductance that exceeds the sum of the conductances of two individual dislocations. We shall point out below that such nonlinear behavior may be a feature of tunneling currents associated with dislocations. The same work also reports on measurements of $1/f$ noise. The noise current was found to be proportional to the leakage current, leading to the conclusion that the dislocations act only as indirect noise sources by increasing the leakage current.

3. CALCULATION OF DISLOCATION PROPERTIES

We wish to distinguish between dislocations that intersect the junction and those that do not. We shall refer to intersecting dislocations as *threading dislocations*. Threading dislocations eventually intersect the surface of the device, and therefore are observable by etch pits. Such dislocations have a number of sources. They can originate in the substrate and propagate through the active material during growth; they can be introduced by plastic deformation of the material; or they can originate from lattice or thermal mismatch in the structure.

3.1 DISLOCATION CORES

Classical elasticity theory works well for the description of long-range strain fields of dislocations in real solids, but fails when it comes to a description of the dislocation core. The description of dislocation cores necessarily depends on the microscopic properties of the

materials. If we view the 60° dislocation as the removal of three $\{112\}$ -type half-planes of atoms, followed by the displacement and rejoining of the crystal across the missing plane, the line that describes the termination of the missing plane is the dislocation line. The dislocation core is the cylindrical volume centered at the dislocation line; in that core, bonds are broken, displacements of atoms are large, and displacements cannot be described by linear elastic theory. Because broken bonds are present in the cores, localized states can be introduced in the band gap. Although reconstruction may take place to move these states out of the gap, it is not clear whether this occurs in MCT.

For the undissociated 60° dislocation in MCT, the local states of dislocation cores will depend on the plane on which the extra half-planes of atoms terminate (i.e., whether they are cation or anion terminated), and whether the dislocation belongs to the shuffle or glide set. Dislocations of the glide set in the group IV and III-V semiconductors are frequently found to be dissociated into two Shockley partials separated by a stacking fault. One study on GaAs found that as many as 80% of the 60° dislocations in plastically deformed GaAs were dissociated⁷. It is not clear whether the majority of the 60° dislocations in MCT are dissociated, but is likely that they are. Since the primary effects of the space charge and piezoelectric fields considered here occur at the longer range, where classical elasticity is valid, the effects considered in this work differ little between an undissociated dislocation and a pair of partials.

3.2 CHARGED DISLOCATIONS

The Coulomb fields associated with electrically charged dislocation cores can act as scattering and recombination centers and may attract or repel charged point defects. The presence and amount of charge on a dislocation core depends on the distribution of localized levels in the bandgap in the core and on the Fermi level in the semiconductor. Once the core structure of the dislocation has been determined, we can take a phenomenological view and consider only the electric field and potential generated by a given core charge. The effective radius of the field is taken to be that distance from the dislocation core within which the field is large enough to move a minority carrier to the core in a carrier lifetime as measured in the dislocation-free material. In this way we define a volume within which the lifetime and mobility are affected by the presence of a charged dislocation. Only if the affected volume surrounding all the dislocations is a significant fraction of the total volume can we expect their presence to affect lifetime and mobility and hence R_0A . This then relates the depletion radius R or screening radius ℓ to a threshold of the dislocation density n_D .

We represent the dislocation core by a cylindrical charge density $-p$ in a cylinder of radius a in n -type material with donor density n . The potential distribution in the depletion approximation is given by the Poisson equation with charge density p :

$$\begin{aligned}
\rho &= (-p + n)e & r < a \\
&= ne & a < r < R \\
&= 0 & R < r
\end{aligned}$$

where R is the depletion radius, related to a by the charge neutrality requirement $pa^2 = nR^2$. By integrating the Poisson equation we find for the potential

$$\phi(r) = \phi(a) + \frac{en}{2\epsilon} \left[R^2 \ln \frac{r}{a} - \frac{1}{2}(r^2 - a^2) \right]$$

and taking the potential in the bulk ($r \geq R$) as reference, we can evaluate the potential at the core radius

$$\phi(a) = \frac{en}{2\epsilon} \left[\frac{1}{2}(R^2 - a^2) - R^2 \ln \frac{R}{a} \right].$$

Using a dislocation core radius of 5×10^{-8} cm and assuming the bands bend by the bandgap—that is, by a potential (relative to the bulk) at the core radius of 0.1 V for 8- to 12- μ m material, we have calculated R as a function of carrier density, as shown in Figure 2. It turns out that the radius is close enough to the Debye length to make the depletion approximation questionable.

The exact space charge equation is

$$\frac{1}{r} \frac{d}{dr} r \frac{d\phi}{dr} = \frac{2}{L^2} [\sinh(\phi - \phi_F) - \sinh \phi_F]$$

where

$$\phi = \frac{e\phi}{kT}, \quad L^2 = \frac{\epsilon kT}{e^2 n_i}, \quad \phi_F = \frac{E_F - E_i}{kT} = \sinh^{-1} \frac{N_D - N_A}{n_i}.$$

For ϕ small this is

$$\frac{1}{r} \frac{d}{dr} r \frac{d\phi}{dr} \cong \left(\frac{2}{L_D^2} \cosh \phi_F \right) \phi = \frac{2}{L_D^2} \phi$$

to determine C. The effective scattering radius of the dislocation can be obtained from the radius at which the field is reduced to the threshold value discussed above, here denoted E.

$$E = \frac{d\phi}{dr} = \frac{kT}{e} \frac{d\phi}{dr} = \frac{kT}{eL_D} \frac{d\phi}{dx} = \sqrt{\frac{nkT}{\epsilon}} CK_1(x)$$

where n is the carrier density. Here the argument of the Bessel function corresponds to the screening radius, so x is large, and we can use

$$K_1(x) \cong \sqrt{\frac{\pi}{2x}} e^{-x}$$

Finally we obtain the screening radius $\ell = L_D x$ from the solution of

$$\sqrt{\frac{nkT}{\epsilon}} C \sqrt{\frac{\pi}{2x}} e^{-x} = E.$$

The threshold field E is now determined, as discussed above, from the mobility μ and lifetime τ in dislocation-free material

$$E = \ell/\mu\tau = L_D x/\mu\tau.$$

This is substituted in the immediately preceding expression, and the resulting expression solved for the normalized screening radius x. The screening radius ℓ as a function of n is plotted for T = 40 K, 78 K, and 300 K in Figure 2. In the calculation we have used the exact values of the Bessel functions rather than the approximations show above, and the following values of mobility and lifetime:

T, K	μ , cm ² /Vs	τ , s
40	1.9×10^5	1×10^{-6}
78	9×10^4	3×10^{-7}
300	1×10^4	1×10^{-7}

As expected, ℓ turns out to be larger than the depletion radius, confirming the need for the more accurate screening theory. The figure shows the results of both calculations: the radius R of the field region surrounding the positively charged core as a function of carrier density calculated on the basis of the depletion approximation, and the screening radius ℓ of the region inside which the electric field due to the charged core exceeds the critical value. The threshold is consistent with the experimental observations^{1,4} that the R_0A values and minority-carrier lifetimes decrease rapidly with the dislocation density n_D when it exceeds 10^6 cm⁻². The

calculation of the screening radius ℓ is more reliable than the depletion approximation, since it takes into account the thermal excitations of the carriers. With the parameters we use, the calculation of the screening radius over five orders of magnitude of the carrier density, and over the temperature range from 40 to 300 K, results in values of ℓ that range from about $5 L_D$ at $n = 10^{10} \text{ cm}^{-3}$ to $10 L_D$ at $n = 10^{15} \text{ cm}^{-3}$, resulting in an overall rate of decrease of ℓ with carrier density at a rate slightly slower than $L_D^{-1/2}$.

The carrier density in the neutral region of the extrinsic device material generally equals or exceeds 10^{14} cm^{-3} . The screening length at this carrier density, even at room temperature, is less than $5 \mu\text{m}$, and it is only about $3 \mu\text{m}$ at 40 K. Since the mobile carrier density in the neutral regions generally exceeds 10^{14} cm^{-3} , we can conclude that in the presence of mobile carriers, especially at cryogenic temperatures, the core charge fields have a limited range. This is also consistent with the observations of Johnson *et al.*¹ that the responsivity of p on n heterojunction devices (dependent on diffusion in the neutral region) is little affected even by high dislocation densities.

The situation is less clear in the vicinity of the metallurgical junction of the p-n diode detector. In this region, the carrier density is reduced to the intrinsic value n_i under zero bias, and below that with reverse bias. We estimate a value of n_i about $5 \times 10^{11} \text{ cm}^{-3}$ for $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$ at 40 K, with a decrease of about a factor of 2 for every 5 mV of reverse bias caused by the separation of the quasi-Fermi levels. This reduction of the carrier density is confined to a small region, and an assessment of its quantitative effect will require detailed computation. Nonetheless, it is worth observing that we estimate a screening radius of $27 \mu\text{m}$ for this carrier density. This is consistent with the experimental finding⁴ that R_0A at this temperature is already decreasing with dislocation densities in the 10^5 cm^{-2} range, corresponding to this order of separation between dislocation lines. The literature values of $n_i = 1.2 \times 10^{14} \text{ cm}^{-3}$ at 77 K and $n_i = 4.0 \times 10^{16} \text{ cm}^{-3}$ at 300 K⁷ correspond to progressively shorter screening lengths, which is also consistent with the experimental observation that the decrease in R_0A with dislocation density has a higher threshold at the higher temperature.

The transition from the neutral to the depleted region is not abrupt, so that the screening is a function of both the distance from the dislocation core and the distance from the junction plane. For a dislocation perpendicular to the junction plane, the geometry has cylindrical symmetry, making it effectively two-dimensional. The geometry is more complicated for a dislocation inclined relative to the junction plane.

Because its form is geometrically simpler and more readily visualizable in this context, we have discussed screening entirely in connection with the effect of core charges. It should be kept in mind that similar, although quantitatively different, considerations apply to the screening of the piezoelectrically induced potentials, to be discussed next.

3.3 DISLOCATION STRAIN FIELDS AND PIEZOELECTRIC POTENTIALS

The MCT crystal structure lacks inversion symmetry and is therefore piezoelectric. In a piezoelectric crystal the strain field around a dislocation may generate an electric field. Conversely, an electric field applied to a piezoelectric crystal can produce a stress that may deform the crystal. This self-consistent field problem has been treated rigorously for a straight dislocation in a homogeneous insulating crystal by Saada⁸. The general analysis was extended by Faivre and Saada⁹ to include the effects of screening by free carriers. Several qualitatively significant conclusions can be drawn from this work:

1. The strains are only slightly affected by the electric field; that is, the inverse piezoelectric effect is small, and the electric field can be calculated directly without imposing the requirement of self-consistency.
2. The electric field in the vicinity of the dislocation core can be quite large, and the dislocation can interact strongly with charges (point defects or mobile carriers) in its vicinity.
3. In a semiconductor, the dislocation preferentially traps carriers of a particular sign.
4. The estimated screening distance is of the order of several Debye lengths.

Booyens and Vermaak¹⁰ have applied Saada's analysis to $\langle 110 \rangle / (111)$ dislocations in zinc blende III-V semiconductors and Booyens and Basson¹¹ to CdTe and MCT. They do not explicitly include the effects of screening as analyzed by Faivre and Saada⁸ in these calculations. As a result, they find that an uncharged dislocation, in the absence of screening, generates a net radial field. They then suggest that the dislocation core is likely to acquire a charge that compensates for the radial field and so minimizes the total energy of the dislocation.

We have extended the analysis to the commonly observed 60° dislocations in intrinsic MCT. The dislocation lies along $[101]$ and has a Burgers vector \mathbf{b} at 60° to the dislocation line ξ . x_1 is chosen parallel to the edge component of the Burgers vector \mathbf{b} ; x_2 is chosen as the normal to the slip plane; and x_3 is chosen parallel to ξ . The transformation matrix from principal axis coordinates is

$$T_{ij} = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 & -2 & 1 \\ \sqrt{2} & \sqrt{2} & \sqrt{2} \\ -\sqrt{3} & 0 & \sqrt{3} \end{pmatrix},$$

and thus the transformation of the piezoelectric tensor is

$$d'_{ijk} = T_{ip} T_{jq} T_{kr} d_{pqr}.$$

In cubic symmetry and principal axis coordinates, all the nonzero components of the piezoelectric tensor are equal: $d_{123}=d_{132}=d_{231}=d_{213}=d_{312}=d_{321}=d$. The nonzero components of the transformed tensor are

$$\begin{aligned} d'_{133} &= d'_{313} = d'_{331} = -d'_{111} = \frac{2}{\sqrt{6}}d \\ d'_{112} &= d'_{121} = d'_{211} = d'_{233} = d'_{323} = d'_{332} = \frac{1}{\sqrt{3}}d \\ d'_{222} &= \frac{2}{\sqrt{3}}d; \end{aligned} \quad (1)$$

or, using two-index (Voigt) notation,

$$(d') = \begin{pmatrix} -\frac{2}{\sqrt{6}}d & 0 & \frac{2}{\sqrt{6}}d & 0 & 0 & -\frac{1}{\sqrt{3}}d \\ -\frac{1}{\sqrt{3}}d & \frac{2}{\sqrt{3}}d & -\frac{1}{\sqrt{3}}d & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{\sqrt{3}}d & \frac{2}{\sqrt{6}}d & 0 \end{pmatrix}.$$

The piezoelectrically induced bound charge density is¹²

$$\rho(r) = \frac{1}{\epsilon} d'_{ijk} e_{jk,i} \text{ with } e_{jk,i} \equiv \frac{\partial}{\partial x_i} e_{jk} \quad (2)$$

where ϵ is the isotropic dielectric constant, e_{jk} is the jk component of the strain tensor, and summation over repeated indexes is implied.

For a screw dislocation in linear anisotropic theory, the displacements are

$$u_3 = -\frac{b_s}{2\pi} \tan^{-1} A \left(\frac{x_2}{x_1} \right),$$

$$u_1 = u_2 = 0,$$

where b_s is the magnitude of the screw component of the Burgers vector and

$$A = \frac{\sqrt{c'_{44}c'_{55}}}{c'_{44}}; \quad c'_{44} = c_{44} - \frac{1}{3}H; \quad c'_{55} = c_{44} - \frac{1}{6}H; \quad H = 2c_{44} + c_{12} - c_{11}.$$

Here c_{ij} and c'_{ij} are the components of the elastic tensor in the principal axis and transformed coordinate systems respectively. The nonzero components of the strain tensor are [using $e_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i})$]

$$e_{13} = e_{31} \text{ and } e_{23} = e_{32}.$$

Using Eqs. 1 and 2 shows immediately that for this case $r = 0$. Even in anisotropic linear elasticity (assuming the dielectric constant isotropic) the screw component of the 60° dislocation induces no piezoelectric effect. Therefore, pure screw $\langle 110 \rangle$ dislocations will also show no piezoelectric activity even if they are dissociated. We now proceed to consider the edge component, which we will simply denote b . In order to arrive at a transparent expression, we will make the isotropic elastic approximation. This will not affect the order of magnitude calculations that follow.

For an edge dislocation, with ν Poisson's ratio

$$\begin{aligned} u_1 &= \frac{b}{2\pi} \left[\tan^{-1} \left(\frac{x_2}{x_1} \right) + \frac{x_1 x_2}{2(1-\nu)(x_1^2 + x_2^2)} \right]; \\ u_2 &= -\frac{b}{2\pi} \left[\frac{1-2\nu}{4(1-\nu)} \ln(x_1^2 + x_2^2) + \frac{x_1^2 - x_2^2}{4(1-\nu)(x_1^2 + x_2^2)} \right]; \\ u_3 &= 0. \end{aligned}$$

The nonzero strain components are

$$\begin{aligned} e_{11} &= \frac{b}{2\pi} \frac{(1-2\nu)x_2^3 + (3-2\nu)x_1^2 x_2}{2(\nu-1)(x_1^2 + x_2^2)^2} \\ e_{22} &= \frac{b}{2\pi} \frac{(1-2\nu)x_2^3 - (1+2\nu)x_1^2 x_2}{2(\nu-1)(x_1^2 + x_2^2)^2} \\ e_{12} &= \frac{b}{2\pi} \frac{x_1 x_2^2 - x_1^3}{2(\nu-1)(x_1^2 + x_2^2)^2}. \end{aligned}$$

From Eqs. 2 and 1,

$$\rho(r) = \frac{1}{\epsilon} \left(d'_{111} e_{11,1} + 2d'_{112} e_{12,1} + d'_{211} e_{11,2} + d'_{222} e_{22,2} \right)$$

so we need the derivatives

$$e_{11,1} = \frac{b}{2\pi} \frac{(2v+1)x_1x_2^3 + (2v-3)x_1^3x_2}{(v-1)(x_1^2+x_2^2)^3}$$

$$e_{11,2} = \frac{b}{2\pi} \frac{(2v-1)x_2^4 - 6x_1^2x_2^2 + (3-2v)x_1^4}{2(v-1)(x_1^2+x_2^2)^3}$$

$$e_{22,2} = \frac{b}{2\pi} \frac{(2v-1)x_2^4 + 6x_1^2x_2^2 - (2v+1)x_1^4}{2(v-1)(x_1^2+x_2^2)^3}$$

$$e_{12,1} = \frac{b}{2\pi} \frac{x_2^4 - 6x_1^2x_2^2 + x_1^4}{2(v-1)(x_1^2+x_2^2)^3}$$

giving

$$\begin{aligned} \rho(x_1, x_2) &= \frac{d}{\sqrt{3}\epsilon} (-\sqrt{2} e_{11,1} - 2e_{12,1} - e_{11,2} + 2e_{22,2}) \\ &= \frac{bd}{2\pi\sqrt{3}\epsilon} \frac{(2v-3)x_2^4 - 2\sqrt{2}(2v+1)x_1x_2^3 + 30x_1^2x_2^2 + 2\sqrt{2}(3-2v)x_1^3x_2 - (2v+7)x_1^4}{2(v-1)(x_1^2+x_2^2)^3} \end{aligned}$$

When this expression is converted to polar coordinates in the x_1x_2 plane, it becomes somewhat simpler:

$$\rho = \frac{bd}{2\pi\sqrt{3}\epsilon} (\rho_1 + \rho_2 + \rho_3 + \rho_4 + \rho_5)$$

with

$$\rho_1 = \frac{(2v-3)}{2(v-1)} \frac{y^4}{r^6} = \frac{(2v-3)}{2(v-1)} \frac{\cos 4\theta - 4\cos 2\theta + 3}{8r^2}$$

$$\rho_2 = \frac{-2\sqrt{2}(2v-1)}{2(v-1)} \frac{xy^3}{r^6} = \frac{-2\sqrt{2}(2v-1)}{2(v-1)} \frac{2\sin 2\theta - \sin 4\theta}{8r^2}$$

$$\rho_3 = \frac{30}{2(v-1)} \frac{x^2y^2}{r^6} = \frac{30}{2(v-1)} \frac{1 - \cos 4\theta}{8r^2}$$

$$\rho_4 = \frac{2\sqrt{2}(3-2\nu)}{2(\nu-1)} \frac{x^3 y}{r^6} = \frac{2\sqrt{2}(3-2\nu)}{2(\nu-1)} \frac{\sin 4\theta + 2\sin 2\theta}{8r^2}$$

$$\rho_5 = \frac{-(2\nu+7)}{2(\nu-1)} \frac{x^4}{r^6} = \frac{-(2\nu+7)}{2(\nu-1)} \frac{\cos 4\theta + 4\cos 2\theta + 3}{8r^2}$$

so

$$\rho = \frac{bd}{4\pi\sqrt{3}\epsilon(1-\nu)} \frac{1}{r^2} [5\cos 4\theta - \sqrt{2}\sin 4\theta + 2(1+\nu)\cos 2\theta - \sqrt{2}(1-2\nu)\sin 2\theta]$$

$$\equiv \frac{C}{r^2} \sum_{n=2,4} [a_n \cos(n\theta) + b_n \sin(n\theta)]$$

which defines a useful constant C, and four parameters depending only on Poisson's ratio. This simple expression reveals the charge density to have an inverse square decay and quadrupole and hexadecapole terms only. Note in particular that there is no monopole term and hence no net charge induced on the dislocation line. This rather transparent expression has not been revealed in previous work^{10,11}.

To obtain the piezoelectric potential we need to solve the Poisson equation

$$\nabla^2 \Phi = -\frac{4\pi\rho}{\epsilon}$$

with this charge distribution as the source. In two dimensions, the formal solution of the Poisson equation is

$$\Phi(r, \theta) = \frac{1}{\pi} \int_0^\infty r' dr' \int_0^{2\pi} d\theta' \left[\ln \frac{1}{r_>} + \sum_{m=1}^\infty \frac{1}{m} \left(\frac{r_<}{r_>} \right)^m \cos m(\theta - \theta') \right] \rho(r')$$

$$= \frac{C}{\pi} \sum_{m=1}^\infty \frac{1}{m} \int_0^{2\pi} d\theta' [\sin(m\theta)\sin(m\theta') + \cos(m\theta)\cos(m\theta')] \times$$

$$\times \sum_{n=2,4} [a_n \cos(n\theta') + b_n \sin(n\theta')] \int_0^\infty \frac{r' dr'}{r'^2} \left(\frac{r_<}{r_>} \right)^m$$

where $r_>$ and $r_<$ are the greater and lesser, respectively, of r and r' .

In analogy with the solution for the elastic field, the integrand diverges at the origin and is also not normalizable. To deal with this difficulty, we adopt the same procedure that is used to

avoid the divergence of the elastic energy of a dislocation at the origin in continuum theory. We set cutoffs at an inner radius, R_1 , and an outer radius, R_2 . The cutoff at R_1 removes the core singularity and the cutoff at R_2 allows ρ to be expanded in multipoles. Later, we will identify R_1 as the core radius and we will let $R_2 \rightarrow \infty$. We now have

$$\begin{aligned}\Phi(r, \theta) &= \frac{C}{\pi} \sum_{n=2,4} \frac{\pi}{n} \left[a_n \cos(n\theta) + b_n \sin(n\theta) \left(\frac{1}{r^n} \int_{R_1}^r (r')^{n-1} dr' + r^n \int_r^{R_2} (r')^{-n-1} dr' \right) \right] \\ &= C \left\{ \frac{1}{4} [a_4 \cos(4\theta) + b_4 \sin(4\theta)] \left[\frac{1}{2} - \frac{1}{4} \left(\frac{R_1}{r} \right)^4 - \frac{1}{4} \left(\frac{r}{R_2} \right)^4 \right] \right. \\ &\quad \left. + \frac{1}{2} [a_2 \cos(2\theta) + b_2 \sin(2\theta)] \left[1 - \frac{1}{2} \left(\frac{R_1}{r} \right)^2 - \frac{1}{2} \left(\frac{r}{R_2} \right)^2 \right] \right\}\end{aligned}$$

for $r < R_2$. For $r > R_2$

$$\begin{aligned}\Phi(r, \theta) &= C \left\{ \frac{1}{4} [a_4 \cos(4\theta) + b_4 \sin(4\theta)] \frac{1}{4} \left[\left(\frac{R_2}{r} \right)^4 - \left(\frac{R_1}{r} \right)^4 \right] \right. \\ &\quad \left. + \frac{1}{2} [a_2 \cos(2\theta) + b_2 \sin(2\theta)] \frac{1}{2} \left[\left(\frac{R_2}{r} \right)^2 - \left(\frac{R_1}{r} \right)^2 \right] \right\}\end{aligned}$$

which is a multipole potential. The two solutions are matched at R_2 . We can now let $R_2 \rightarrow \infty$ and neglect the terms in R_1 , which amount to a small correction near the core, to get a potential valid for $r \gg R_1$

$$\Phi(r, \theta) = C \left\{ \frac{1}{8} [a_4 \cos(4\theta) + b_4 \sin(4\theta)] + \frac{1}{2} [a_2 \cos(2\theta) + b_2 \sin(2\theta)] \right\}$$

which is independent of r . This rather surprising behavior is a consequence of the inverse square decay of the charge density. In this connection, we can pursue the analogy with the well-known elastic solution for an edge dislocation by pointing out that while the elastic energy per unit length is given by (μ is the isotropic shear modulus)

$$\frac{\mu b^2}{4\pi(1-\nu)} \left(\ln \frac{R_2}{R_1} - 0.5 \right)$$

the piezoelectric energy per unit length is (where C was defined earlier, and taking $\nu = 0.3$)

$$\frac{9\pi^2 C^2}{1.6} \left(\ln \frac{R_2}{R_1} - 0.125 \right)$$

which is six orders of magnitude smaller—justifying our neglect of the reverse piezoelectric effect. In both cases, the first term comes from the energy in an infinite cylinder integrated to R_2 and which diverges logarithmically, and the second term is the effect of the cylindrical surface bounding this dislocation in a finite crystal (the image term).

For a good numerical calculation, we would need the piezoelectric constant of the alloy, which has not been determined. For an order-of-magnitude estimate, we can use the piezoelectric constant of CdTe, $3 \times 10^{-6} \text{ C cm}^{-2}$. The Burgers vector is $4 \times 10^{-8} \text{ cm}$, and Poisson's ratio can be taken as 0.3. Using these values we have computed the charge, potential, and field, shown in Figure 3.

It should be understood that because of the symmetry of the strain field associated with a straight dislocation, several of the possible sources of the piezoelectric potential are identically zero. In the reduced symmetry of curved dislocations (e.g., misfits), kinks, and jogs, these components of the potential do not vanish and may be significant.

3.4 AVERAGING NONLINEAR EFFECTS

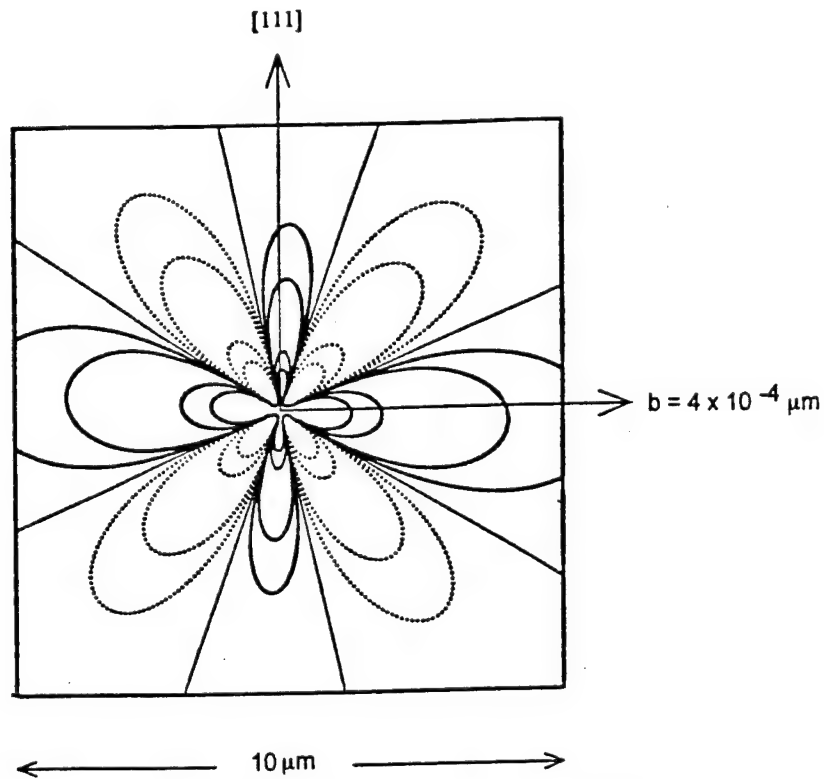
The results we have obtained afford a qualitative insight into the effect of the piezoelectric potential on the characteristics of a pn junction threaded by a dislocation. The angular variation of the potential will alternately add to and subtract from the built-in potential of the junction, so that the barrier seen by a carrier crossing the junction will depend on its trajectory. Because of the strongly nonlinear dependence of both diffusion and tunneling currents on the barrier, the effect of such a barrier variation does not average to zero and the currents can be enhanced by a large factor. To illustrate the principle, we choose as an example the average tunneling probability for carriers crossing a narrow semiconductor junction.

The band-to-band tunneling probability has the form

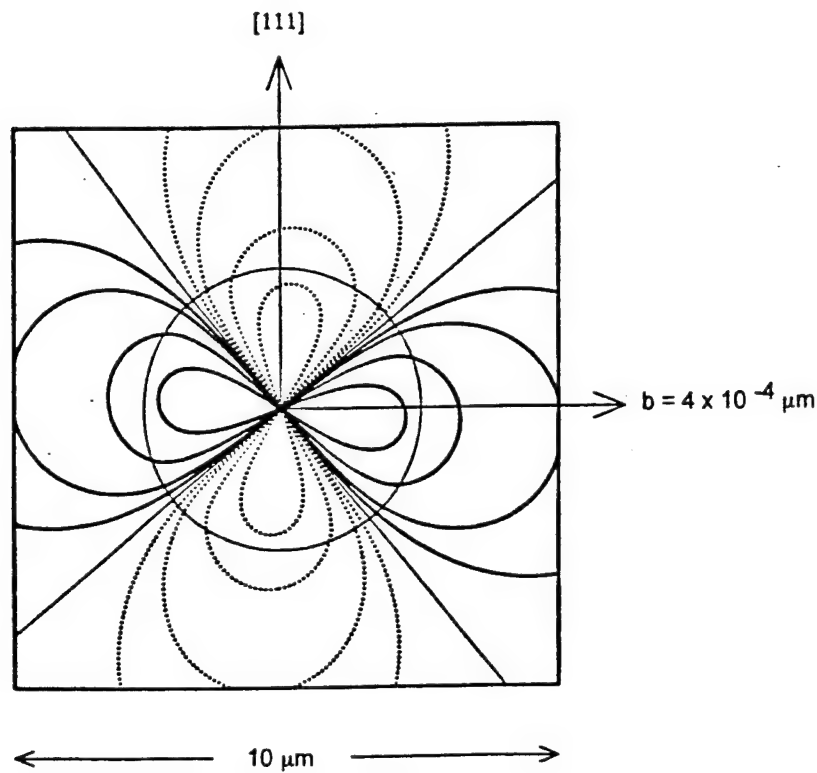
$$p = \exp(-CE_g^{3/2})$$

where E_g is the bandgap and C is a constant containing all other material parameters.

We crudely represent the effect of the dislocation potential, which makes the tunneling barrier a function of the carrier's trajectory, as an angular variation of the bandgap of the form $E_g = E_{g0}(1 + \Delta \sin \theta)$ (it will be apparent presently that the detailed form of the angular variation is irrelevant for this argument). It is likely that this representation of the piezoelectric potential introduced by a dislocation can probably be justified by a WKB calculation.

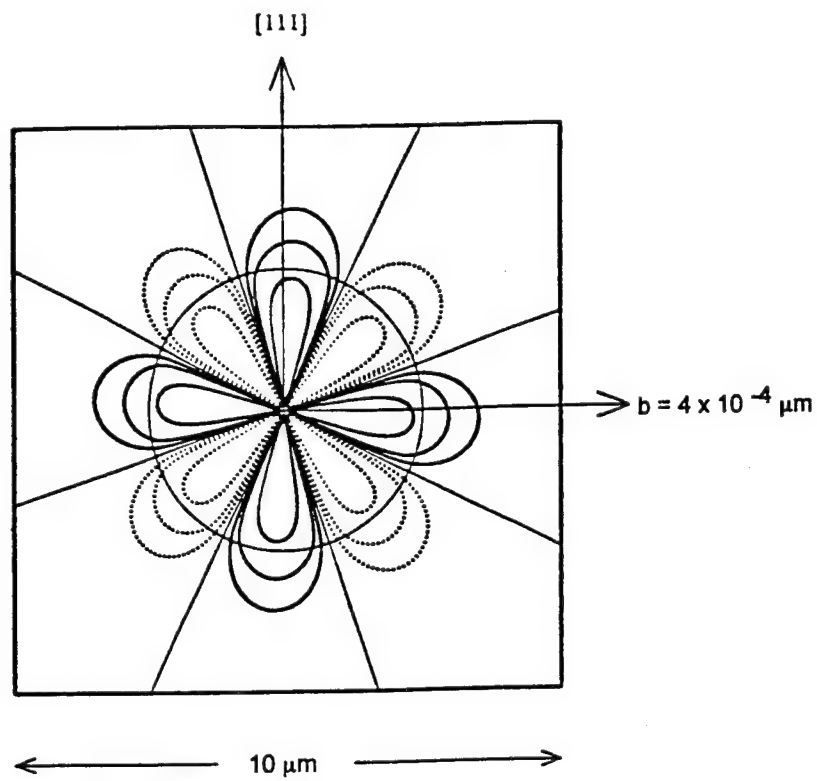


(a) Piezoelectric charge normal to the dislocation line. Solid contours are 1, 0.5, 0.1, and $0.05 \times 10^{12} \text{ e/cm}^3$ going out radially; dotted contours are the negative of these.

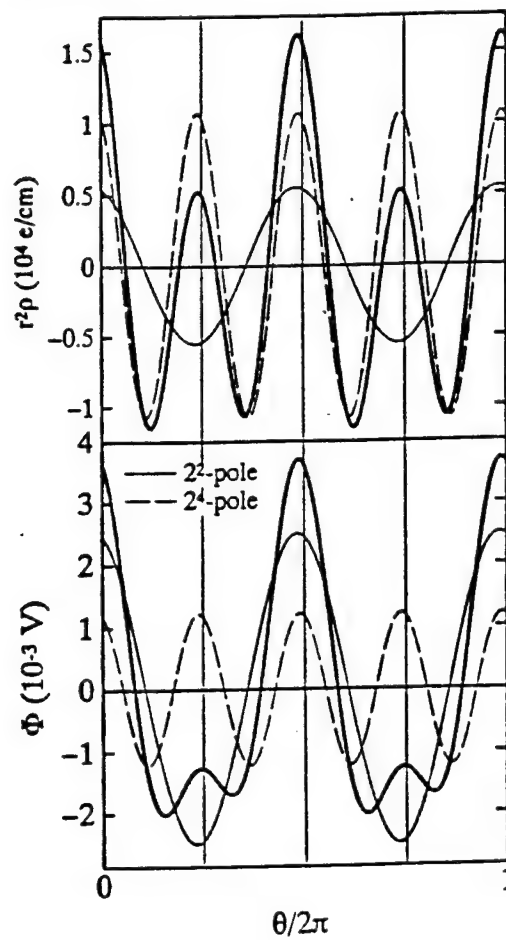


(b) Quadrupole component of the piezoelectric potential.

Figure 3. Piezoelectric properties of a 60° dislocation in $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$

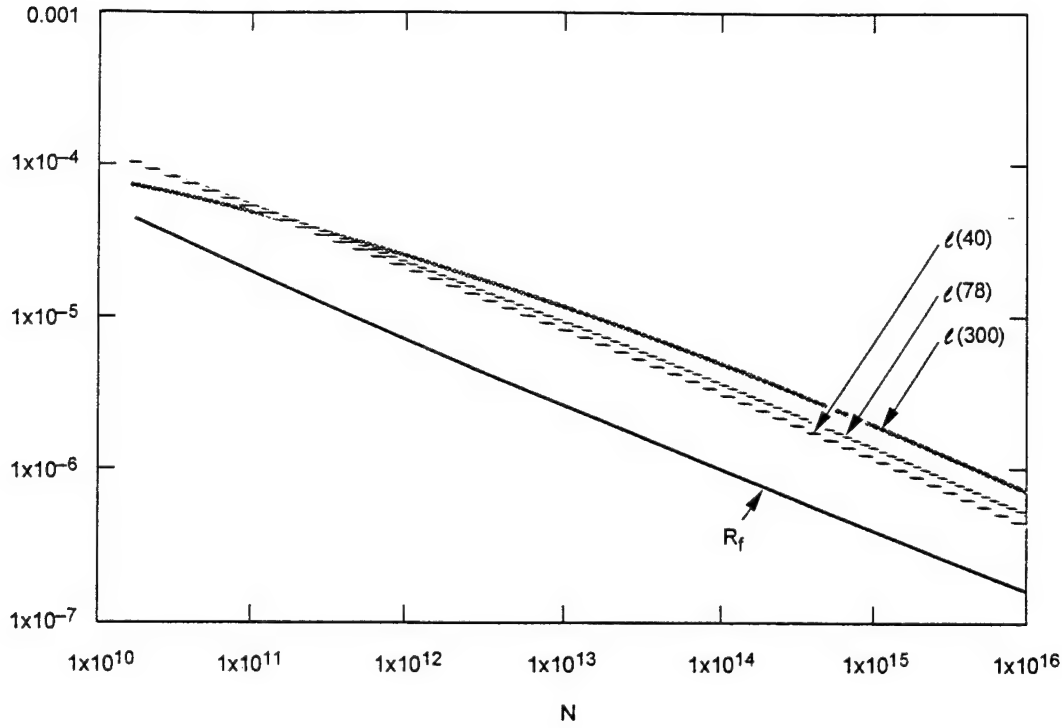


(c) Hexadecapole component of the piezoelectric potential



(d) Angular variation of the piezoelectric charge and potential normal to the dislocation line

Figure 3 (continued). Piezoelectric properties of a 60° dislocation in $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$



mu-072893-jd

Figure 2. Depletion radius R and screening radius ℓ (meters) at 40, 78, and 300 K as functions of carrier density in cm^{-3}

with the extrinsic Debye length

$$L_D^2 = \frac{L^2}{\cosh \phi_F} \cong \frac{\epsilon k T}{e^2 N_D} = \frac{2.77 \times 10^5}{N_D}, \quad N_D \text{ in } \text{cm}^{-3}.$$

Then with $x = \frac{r}{L_D}$ we have $\frac{1}{x} \frac{d}{dx} x \frac{d\phi}{dx} = \phi$, the zero-order modified Bessel equation. The solution of interest is the decreasing Hankel function $K_0(x)$.

The depletion approximation should be valid for $r < R - L_D$. This does not leave much room since R and L_D are comparable. One way to proceed is to let $\phi = CK_0(x)$ and to evaluate C from $\phi(a)$ at $x = a/L_D$. Since the value of x here is that at the core radius, x is small and we can use

$$K_0(x) \cong \ln(\gamma x) = 0.116 - \ln x,$$

so

$$K_0\left(\frac{a}{L_D}\right) = 0.116 + \ln\left(\frac{a}{L_D}\right) = \frac{\phi(a)}{C}$$

If Δ is small compared with unity we have

$$E_g^{3/2} \approx E_{g0}^{3/2} \left(1 + \frac{3}{2} \Delta \sin \theta \right)$$

(this approximation is taken purely to simplify the mathematics and has no effect on the argument). We can now compute an average of p

$$\langle p \rangle = \frac{1}{2\pi} \int_0^{2\pi} \exp \left[-CE_{g0}^{3/2} \left(1 + \frac{3}{2} \Delta \sin \theta \right) \right] d\theta = \frac{p_0}{2\pi} \int_0^{2\pi} \exp \left[-\frac{3}{2} CE_{g0}^{3/2} \Delta \sin \theta \right] d\theta = p_0 I_0 \left(\frac{3}{2} CE_{g0}^{3/2} \Delta \right)$$

where $I_0(x)$ is the modified zero-order Bessel function and p_0 is the tunneling probability in the absence of any dislocations. From this last form it can be seen that the average tunneling probability could have been written as

$$\langle p \rangle = p_0 I_0 \left(\frac{3}{2} \Delta \ln | p_0 | \right).$$

It can be seen that although the fractional bandgap "angular modulation" Δ is small, the argument of the Bessel function need not be, and the factor multiplying the unmodulated bandgap transition probability can be quite large. It is also evident that any other angular variation would lead to very similar results. The angular variation Δ can be estimated from the angular potential plots of Figure 3 to be roughly 0.05 to 0.1. The parameters determining p_0 (e.g., effective mass, doping, temperature) cover a wide range. For typical values of these parameters the multiplying Bessel function ranges from near unity to about 10, so $\langle p \rangle / p_0$ is in rough agreement with the observed¹ effect of a single dislocation on R_0A of a pixel.

As another consequence of the type of nonlinearity examined here, the effect of overlapping piezoelectric potentials of neighboring dislocations is expected to be enhanced over what might be expected from a linear superposition. This may be a possible clue to the observed inverse faster than linear falloff of R_0A with dislocation density.

4. CONCLUSIONS

We have evaluated the electric fields associated with dislocation charges and strain fields, and estimated their effects on junction performance. We conclude that the fields associated with core charges alone are not likely to account for the observed degradation of junction performance, particularly when the lower carrier concentration is on the n side of the junctions so

the principal minority carriers are holes. The longer range piezoelectric potentials, which we have calculated explicitly for the common 60° dislocation, are likely to have a larger effect, especially on R_0A through the tunneling leakage current. The nonlinear dependence of the junction characteristics on dislocation density is examined, and is proposed as a natural explanation for the observed nonlinear behavior of R_0A .

REFERENCES

1. S.M. Johnson, R. Righer, J.P. Rosbeck, J.M. Peterson, S.M. Taylor, and M.E. Boyd, *J. Vac. Sci. Technol. B* **10**, 1499 (1992).
2. S.H. Shin, J.M. Arias, D.D. Edwall, M. Zandian, J.G. Pasko, and R.E. DeWames, *J. Vac. Sci. Technol. B* **10**, 1492 (1992).
3. K. Kosai and W.A. Radford, *J. Vac. Sci. Technol. A* **8**, 1252 (1990).
4. R.E. DeWames, private communication (1993).
5. P.I. Baranskii, A.E. Belyaev, O.P. Gorodnicheii, and S.M. Komirenko, *Sov. Phys. Semicond.* **24**, 73 (1990).
6. J.C. Zolper and A.M. Barnett, *IEEE Trans. Elect. Dev.* **37**, 478 (1990).
7. D. Gerthsen, F.A. Ponce, and G.B. Anderson, *Philosophical Magazine A* **59**, 1045 (1989).
8. G. Saada, *Phys. Stat. Sol. (B)* **44**, 717 (1971).
9. G. Faivre and G. Saada, *Phys. Stat. Sol. (B)* **52**, 127 (1972).
10. H. Booyens and J.S. Vermaak, *J. Appl. Phys.* **50**, 4302 (1979).
11. H. Booyens and J.H. Basson, *Phys. Stat. Sol. (A)* **85**, 243 (1984).
12. P. Fedders, *J. Appl. Phys.* **54**, 1804 (1983).

APPENDIX F

Temperature dependence of band gaps in HgCdTe and other semiconductors

S. Krishnamurthy, A.-B. Chen, A. Sher, and M. van Schilfgaarde

J. Electron. Mater. **24**, xxxx (1995)

Temperature Dependence of Band Gaps in HgCdTe and Other Semiconductors

SRINIVASAN KRISHNAMURTHY,* A.-B. CHEN,[†] A. SHERA,* and M. VAN SCHILFGAARDE*

*SRI International, Menlo Park, CA 94025

[†]Physics Department, Auburn University, Auburn, AL 36849

Band-edge shifts induced by the electron-phonon interaction are calculated for HgCdTe alloys and various semiconductor compounds starting from accurate zero-temperature band structures. The calculated temperature variation of gaps agrees with experiments to better than 10% in all materials except InAs and InSb where the deviation is about 50%. While the simple picture that the intra (inter)-band transitions reduce (increase) the gap still holds, we show that both the conduction band edge E_c and valence band edge E_v move down in energy. These shifts in E_v affect the valence band offsets in heterojunctions at finite temperature. The temperature variations of valence band offset and the electron effective mass are also reported.

Key words: III-V semiconductors, band offset, electron-phonon interactions, HgCdTe and alloys, temperature-dependent band structures

INTRODUCTION

The temperature (T) dependence of energy gaps of semiconductors is of great physical and technological interest. The quantities such as band offset and effective mass depend sensitively on the temperature variation of band edges. Numerous theoretical¹⁻⁹ and experimental¹⁰⁻¹⁹ studies have been undertaken to obtain both qualitative and quantitative variations of various gaps in semiconductors. The gap decreases with increasing temperature in medium-gap and wide-gap semiconductors, and it increases in small-gap materials such as HgCdTe, PbS, PbSe, and PbTe. The thermal expansion of the lattice and electron-phonon interactions are usually considered causes for the temperature variation of the band structures. Thermal expansion always reduces gaps.

In a perturbation-theory treatment of electron-phonon interactions, the intraband transitions reduce the gap whereas interband transitions increase it, and the net shift in the gap can be positive or negative. Here we calculate the gap variation in $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ alloys, GaAs, InAs, InSb, InP, and CdTe compounds, starting from accurate band structures, wave functions, proper phonon dispersion relations, and taking account of matrix elements of the electron-phonon interactions. The contributions from each phonon branch to each electron band have been obtained to assist physical understanding of the underlying causes of the variations. We show that both conduction and valence band edges move down in energy. When the valence band edge moves more than the conduction band edge, the gap increases with T , as in the case of some $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ alloys with $x < 0.5$. The reverse occurs for all III-V compounds studied and $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ with $x > 0.5$. This observation has an important effect on our understanding of the variation of band offsets in semiconductor heterojunctions. In addition to the gap, other features of the band structure change with temperature and will affect the spectral variations of the absorption coefficient and transport properties.

METHOD

Our calculation of the temperature dependence of the band gap starts with accurate band structures.

(Received October 4, 1994; revised January 8, 1995)

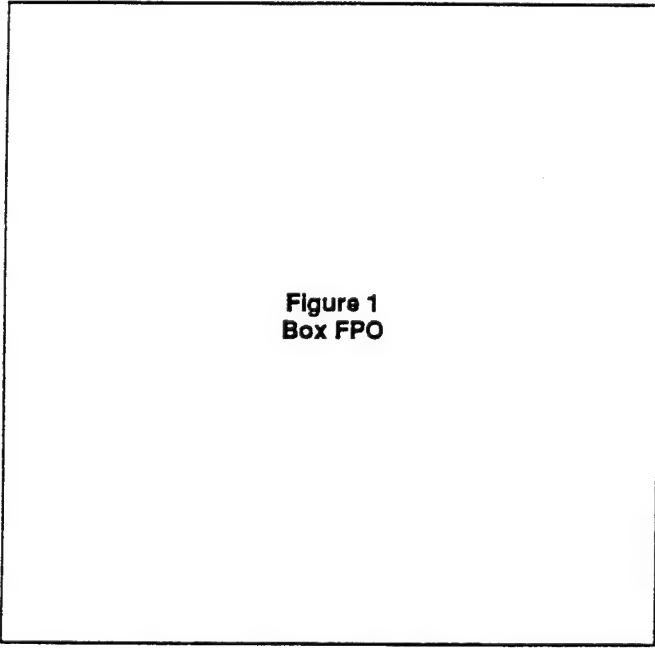


Figure 1
Box FPO

Fig. 1. Change in the band gap of $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ with temperature.

Empirical pseudopotential form factors are used to construct a hybrid pseudopotential tight-binding (HPTB) Hamiltonian. The pseudopotential part of the Hamiltonian is universal—it applies to all group IV, III-V, and II-VI compounds. The smaller tight-binding part is expressed in a minimum set of sp^3 Slater orbitals per atom. This Hamiltonian is then transformed into an orthonormal basis. A site-diagonal spin-orbit Hamiltonian is then added. Parameters in the tight-binding perturbation are chosen to fine-tune the band structures to agree well with experiments.^{20,21} Various results obtained using these band structures are found to be quite reliable.^{21–23} The present study subjects the accuracy of the wave functions as well as the energies to a sensitive test.

The dilation contribution to the band gap reduction is given^{5,9} by $3\alpha_r B \partial E_g / \partial P$, where the thermal expansion coefficient of the lattice α_r , the bulk modulus B , and the change in the gap with pressure are obtained from the literature.¹⁹ The electron-phonon interactions with all phonon branches that cause the band structure changes are treated in perturbation theory. The total Hamiltonian is assumed to be a sum of potentials from single atoms. The atomic potential in the solid is traditionally expanded in a Taylor series, with only the leading term retained, and the energy shifts it causes are evaluated in second-order perturbation theory. However, it has been demonstrated by a number of researchers^{3,4,6} that retention of first-order perturbation terms with a second term in the Taylor series expansion is necessary to preserve symmetry. We retain both terms. The change in the energy at a given wave vector k is

$$\Delta E_{nk} = \langle nk | V_2 | nk \rangle + \sum_{n'k'} \frac{\langle nk | V_1 | n'k' \rangle^2}{E_{nk} - E_{n'k'}} \quad (1)$$

where V_1 and V_2 are the first two terms in the Taylor expansion of the total electron-phonon potential in powers of atomic displacements ξ . In the TB formalism, Eq. (1) can be written in terms of the matrix elements

$$\langle l'j'\alpha | V_1 | lj\alpha \rangle = \nabla V_{\alpha\alpha'}(d_{ll'}^{jj'}) \cdot (\xi_{lj'} - \xi_{lj}), \quad (2)$$

$$\text{and } \langle l'j'\alpha | V_2 | lj\alpha \rangle = \frac{1}{2} [\xi_{lj'} \cdot (\nabla)^2 V_{\alpha\alpha'}(d_{ll'}^{jj'}) \cdot \xi_{lj'} + \xi_{lj} \cdot (\nabla)^2 V_{\alpha\alpha'}(d_{ll'}^{jj'}) \cdot \xi_{lj}], \quad (3)$$

where $d_{ll'}^{jj'}$ is the position vector connecting atomic sites l , species (anion or cation) j and site l' , species j' , and $V_{\alpha\alpha'}(d_{ll'}^{jj'})$ is a HPTB matrix element between the orbitals α and α' located on those atoms. From the quantum theory of harmonic crystals, the atomic displacements ξ can be expressed in terms of normal modes; that is, phonons. We have

$$\xi_{ij} =$$

$$\left[\frac{\hbar}{2NM_j} \right]^{\frac{1}{2}} \sum_q \omega_{\lambda q}^{-\frac{1}{2}} \left[e_{\lambda q}^j a_{\lambda q} e^{iq(1+\tau_j)} + e_{\lambda q}^{*j} a_{\lambda q}^\dagger e^{-iq(1+\tau_j)} \right] \quad (4)$$

where q and ω are phonon wave vector and frequency, λ denotes phonon branch, $a(a^\dagger)$ is a phonon annihilation (creation) operator, M is the atomic mass, and e is an eigenvector in a diamond or zinc-blende structure of the six-dimensional dynamical matrix eigenvalue problem

$$M\omega^2 e = D(q)e \quad (5)$$

Evaluation of the matrix elements given by Eqs. (2) and (3) requires knowledge of spatial variations of the interatomic TB matrix elements $V_{\alpha\alpha'}$. In Harrison's universal TB approach,²⁴ these matrix elements scale as d^{-2} . In our generalization, we assume that $V_{\alpha\alpha'}$ varies as d^{-m} and the repulsive first-neighbor pair energy, following Harrison's overlap argument, as η/d^{2m} . The two unknowns m and η are determined by requiring that the calculated equilibrium bond length and bulk modulus agree well with experiments. This approach, with electrons and phonons treated from the same underlying Hamiltonian, has previously been used successfully to explain hot electron transistor characteristics²³ and is also in fairly good agreement with first-principles calculations.²⁵ The dynamical matrix D is calculated from the valence force field model.²⁶

The calculational procedure is as follows. For a chosen material, m and η are evaluated. Then the first and second derivatives of all interatomic matrix elements are obtained. The dynamical matrix is diagonalized to obtain ω and e as a function of q and λ . The phonon structures and electronic band structures are used [Eqs. (1) through (4)] to obtain the change in the band energy at a given k . The polar coupling terms are

included in the longitudinal optical phonon contributions. When we are interested in studying the change in the direct gap, k is taken to be zero. However, when the temperature variation of the effective masses or indirect gap are studied, non-zero k values must be used and the Brillouin zone summation in Eq. (1) should be carried over the entire zone with reduced, or no, symmetry.

RESULTS

The calculated band-gap change as a function of T in $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ is shown in Fig. 1. With increasing T , the direct gap increases in $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$. Notice that the calculated values are typically within 10 to 15 meV of experimental values.¹⁶⁻¹⁸ The cross (\times) at $T = 0$ represents the calculated zero-point correction to the gap (13.6 meV for $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$). The zero-temperature band gap calculated without electron-phonon interactions should have this correction subtracted for comparison to experimental values.

The change in the gap is traditionally explained in terms of inter- and intraband interactions. The intravalence (conduction) band interactions push the valence (conduction) band edge up (down), thus reducing the gap. Similarly, the valence-conduction band interactions increase the gap. Hence, one might expect the gap to decrease in wide-gap semiconductors and possibly increase in small-gap semiconductors. In addition, arguments based only on total density of states and ignoring variations in matrix elements will predict the valence band edge E_v move up in energy, because the hole effective mass is one to two orders of magnitude larger than the electron mass. As seen from Fig. 2, our detailed calculations of band edge movements in $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ do not support this traditional view. We find that both E_v and E_c (solid lines) move down in energy. This same trend is observed in other semiconductor compounds studied (GaAs, InP, InAs, InSb, and GaSb). The movement of the valence (solid line) and conduction (dashed) band edges due to interaction with other bands is also shown in Fig. 2. The interaction of the band edges

with the conduction bands (CBs) is much stronger than with valence bands (VBs), and consequently both E_v and E_c move down in energy. E_v moves much more than E_c and the gap increases.

We analyze the strength of inter- and intraband electron-phonon interactions by presenting the contributions from each band and from each phonon mode. Table I lists the calculated values for $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ at 300K. Although spin is included in our band structure calculations, only spin averaged values are listed. Band indices 1 to 4 correspond to VBs, and the others to CBs. The first two rows show changes in E_v and E_c due to interactions with various bands. Contributions from each phonon mode are listed in the remaining rows. The lowest VB is about 12 eV below E_v and E_c , and hence the interaction does not affect the band edges. We see that the interaction with other VBs tends to push the band edges up in energy, as expected. Note that the top two valence

Figure 2
Box FPO

Fig. 2. Variation of conduction (dashed line) and valence (solid line) band edges of $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ with temperature.

Table I. Calculated Change in the Valence (v) and Conduction (c) Band Edge Energies (in meV) of $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ Alloy

Band		1	2	3	4	5	6	7	8
Total	v	5.90	13.99	56.85	88.51	-80.26	-92.49	-97.45	-102.07
	c	1.04	2.84	17.27	34.07	-23.81	-25.43	-43.96	-42.48
TA	v	4.03	6.21	30.06	41.52	-49.86	-62.17	-61.04	-49.62
	c	0.26	0.65	12.41	27.47	-3.58	-14.81	-24.50	-15.40
LA	v	0.91	3.18	3.14	10.53	-10.63	-5.92	-11.41	-20.59
	c	0.56	1.29	1.27	1.72	-14.94	-3.00	-2.66	-7.17
LO	v	0.33	2.11	6.12	11.22	-9.52	-4.98	-11.80	-17.40
	c	0.13	0.63	0.75	0.44	-3.78	-2.45	-1.18	-7.31
TO	v	0.63	2.48	17.53	25.24	-10.26	-19.42	-13.20	-14.45
	c	0.10	0.27	2.84	4.44	-1.52	-5.16	-15.62	-12.59

Note: Contributions from interaction with various phonon modes are shown in rows 3 to 10.

Figure 3
Box FPO

Fig. 3. Derivative of direct gap with temperature for various semiconductor compounds and alloys as a function of zero-temperature gap. The vertical lines represent $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$, InSb , InAs , $\text{Hg}_{0.5}\text{Cd}_{0.5}\text{Te}$, GaAs , and CdTe , respectively.

Table II. Calculated E_v [meV], γ [eV], c [eV], and the Effective Mass Ratio of $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ Alloy as Functions of Temperature

T	E_v	γ	c	$m^*(T)/m^*(0)$
1.00	113.60	47.7656	0.0588	1.0000
10.00	112.67	47.7553	0.0588	1.0005
20.00	112.56	47.7169	0.0592	1.0072
30.00	114.44	47.6421	0.0598	1.0199
40.00	117.15	47.5582	0.0607	1.0361
50.00	120.42	47.4461	0.0615	1.0532
60.00	123.96	47.3310	0.0624	1.0714
70.00	127.65	47.2091	0.0634	1.0904
80.00	131.44	47.0821	0.0643	1.1095
90.00	135.28	46.9418	0.0653	1.1288
100.00	139.17	46.7964	0.0662	1.1483
150.00	158.85	46.1544	0.0712	1.2529
200.00	178.73	45.5930	0.0767	1.3669
250.00	198.68	45.2441	0.0832	1.4938
300.00	218.66	45.1167	0.0908	1.6342
350.00	238.65	45.3460	0.1000	1.7913
400.00	258.66	46.0193	0.1115	1.9672
450.00	278.67	47.3751	0.1263	2.1657
500.00	298.69	49.8338	0.1468	2.3919
550.00	318.71	54.0581	0.1763	2.6491
600.00	338.74	61.7125	0.2238	2.9445

Note: $m^*(0)$ is 0.008. The zero point correction is 13.6 meV.

bands contribute the most to the fundamental band-edge changes. However, interaction of the band edges with CBs is even stronger and negative. Particularly, the interaction of E_v with all conduction bands is strong. As E_v moves down more than E_c , the gap increases in $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$. To understand the role of various phonons, we display the contribution from

each mode separately. The phonon-induced changes in the band edges at 300K in $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ are listed in the third through tenth rows of Table I. We see that acoustic phonons account for about 75% of the total change in the valence and conduction band-edge energies. The selection rules wipe out interband matrix elements between electrons and polar longitudinal phonons, so in spite of the stronger coupling constant, they do not dominate this phenomenon as they do with mobilities, which depend on intraband matrix elements.

Our calculations for other compounds show a qualitatively similar role for phonons. In addition to $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$, we studied the band-gap variation with temperature in GaAs , InAs , InP , InSb , GaSb , and CdTe compounds and HgCdTe alloys. The gap changes linearly at high temperatures ($>150\text{K}$). The calculated dE_v/dT values (circle) are compared with experiments (cross) values in Fig. 3. We see that the calculations produced correct trends in all these materials, but compare less favorably with experiments in InAs and InSb . However, it is important to note that the sign of the change is not exclusively determined by the magnitude of the zero-temperature gap. For example, although $\text{Hg}_{0.70}\text{Cd}_{0.30}\text{Te}$ and InSb have the same zero-temperature gap of 0.235 eV, the InSb gap decreases with T , whereas the $\text{Hg}_{0.70}\text{Cd}_{0.30}\text{Te}$ gap increases with T . The combination of gap size, conduction band width, and intervalley separations gives rise to these interesting variations in the gap with T .

The observation that both E_v and E_c move down in energy has an important effect on band offsets in heterojunction-based devices. For example, the zero-temperature valence band offset between $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ and CdTe is believed to be around 350 meV. However, we find that at 300K, E_v in $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ and in CdTe moves down by 215 and 30 meV, respectively. If the dipole contribution remains the same, the valence band offset decreases to 165 meV at 300K. The contention that the dipole contribution is nearly temperature independent stems from the observation that any shift in the average effective crystal potential should effectively be screened out, since these semiconductors are good dielectrics ($\epsilon \geq 10$). The temperature variation of the bands should be taken with respect to a fixed average potential. For our Hamiltonian, the valence band edge movements in each side of the junction are calculated with reference to a fixed average state. Thus, the calculated temperature dependence of the difference in the VB edge of the constituent heterojunction materials effectively governs the temperature dependence of the band offset. In addition to the electron-phonon interactions discussed above, lattice dilation changes the band edges differently.⁹ This effect is not included here. In any case, this band offset change has important implications for the design of abrupt heterojunction infrared (IR) absorption and confined well laser devices.

In principle, the band structure at any wave vector k will change with temperature. With the change in

the fundamental gap, the band curvature (or effective mass) also changes thus affecting the optical and transport properties of the material. The self-energy calculated in this method will include the effect of scattering due to phonons and the change in the temperature-dependent band structure self-consistently. In the case of a narrow-gap material such as $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$, the effective mass alone does not explain the low-energy portion of the conduction band structure. The lowest CB energy at any k is best described by a hyperbola, $(\gamma k^2 + c^2)^{1/2} - c$. The calculated band gap, effective mass, γ , and c as functions of T are given in Table II. The effective mass and c are directly proportional to the gap and hence monotonically increase with T . This is expected from a $k \cdot p$ theory argument, but the magnitudes predicted by the two theories differ. γ decreases slightly at lower temperatures and then starts to increase with T . In a previous publication,²⁷ we had simulated these temperature variations of γ and c by adjusting the Hg concentration in HgCdTe alloys to produce proper gap at each temperature. Those values are in remarkable agreement with the values reported in Table II. We conclude that γ and c (given in Table II) can be interpolated to considerable accuracy for any positive gap in the HgCdTe alloys.

CONCLUSIONS

Although the calculations produced correct trends in all materials, the calculated changes in the band gap of InAs and InSb were about a factor of two smaller than in the experiments. We find that our calculated TA phonon frequencies away from zone center in these compounds were considerably larger than those found in experiments. As noted from Table I, a substantial contribution comes from acoustic phonons. Consequently, our theoretical values of $E_g(T)$ are smaller than in experiments. Better predictability should result from improvement in the dynamical matrix calculated from the underlying Hamiltonian. In addition, at higher temperatures higher-order perturbation terms must be included along with finite-temperature "renormalized" bands rather than the zero-temperature bands. Such renormalization affects the monotonic change in the gap and introduces nonlinear terms.

In summary, we have calculated the temperature variations of band gaps in various semiconductors. A fairly accurate HPTB Hamiltonian is used in the calculation of electron and phonon structures. The calculations explain the increase in the band gap of $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$, and the decrease in the band gap of all III-V compounds studied. We show that acoustic phonons make the major contribution. Contrary to traditional thinking based on total density of states arguments, we find that both the valence and the conduction band edges move down in energy. One

important consequence of this observation will be in the band offsets in semiconductor heterojunction devices. Finally, there is a small and usually negligible zero-point motion contribution to low-temperature band gaps arising from electron-phonon interactions.

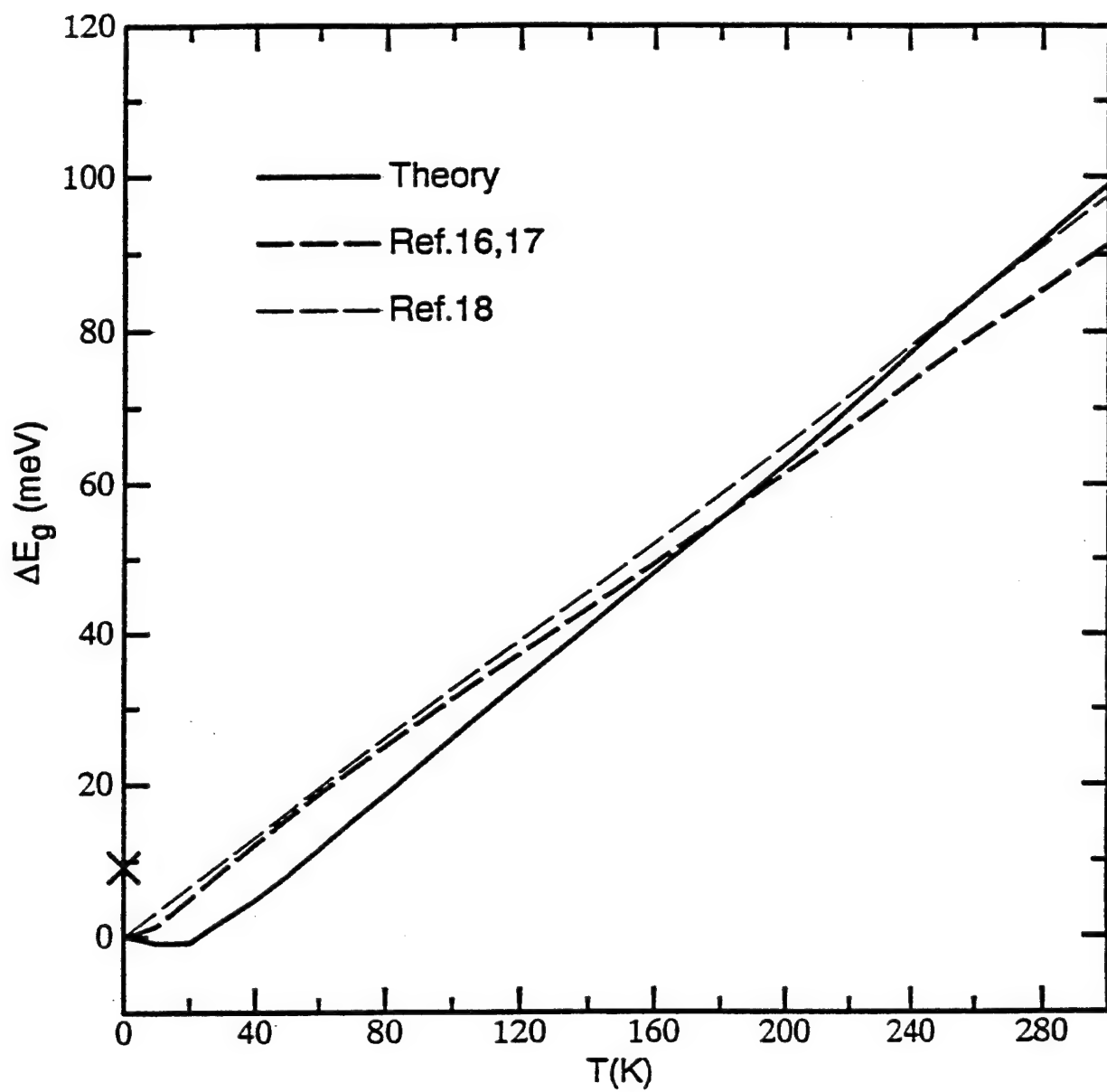
ACKNOWLEDGMENT

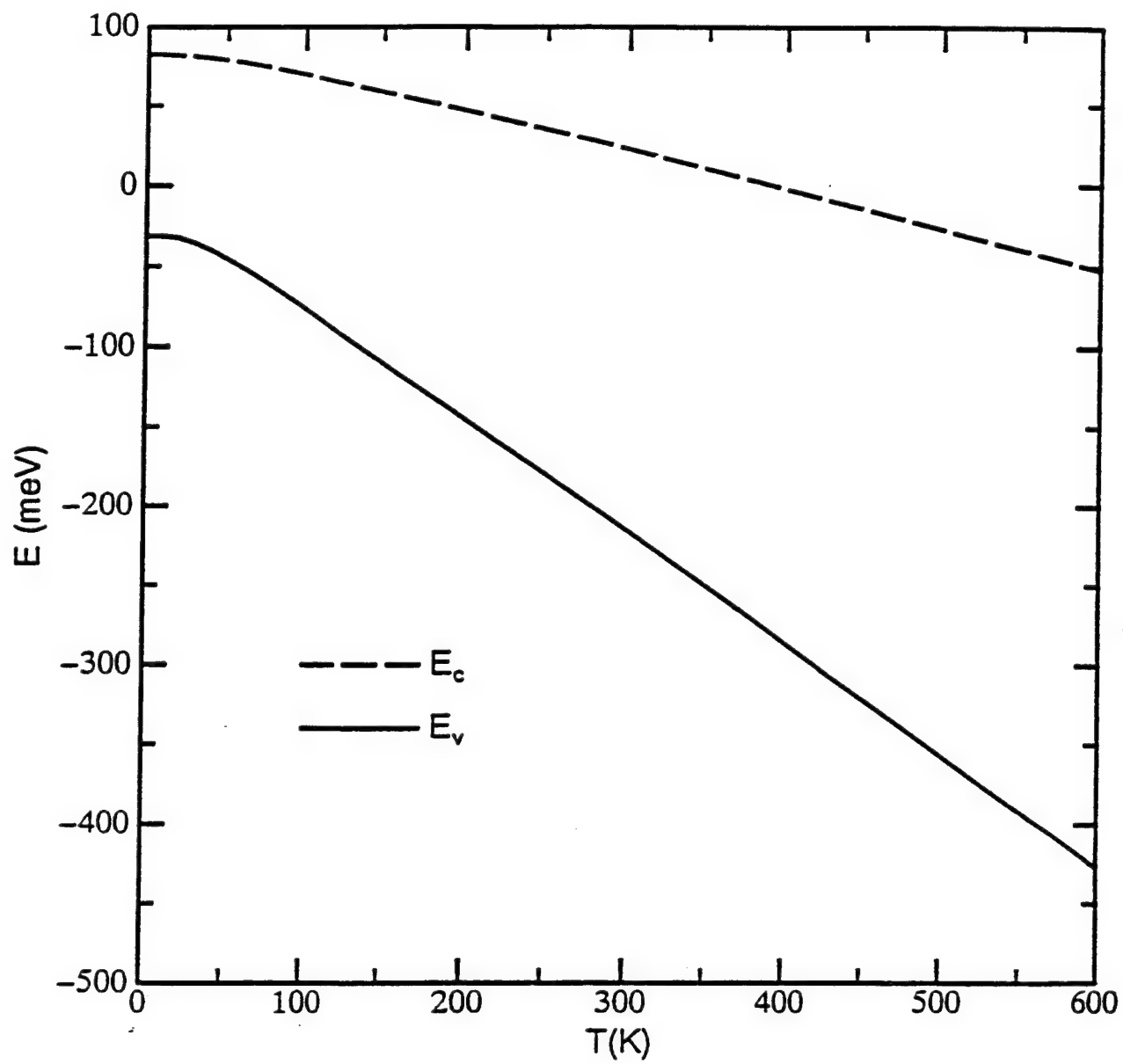
We thank Dr. M. Cardona of the Max Planck Institute, Stuttgart, for pointing us to several references. Funding from ONR (contract N00014-93-C-0091) and ARPA (contract MDA972-92-C-0053) is gratefully acknowledged.

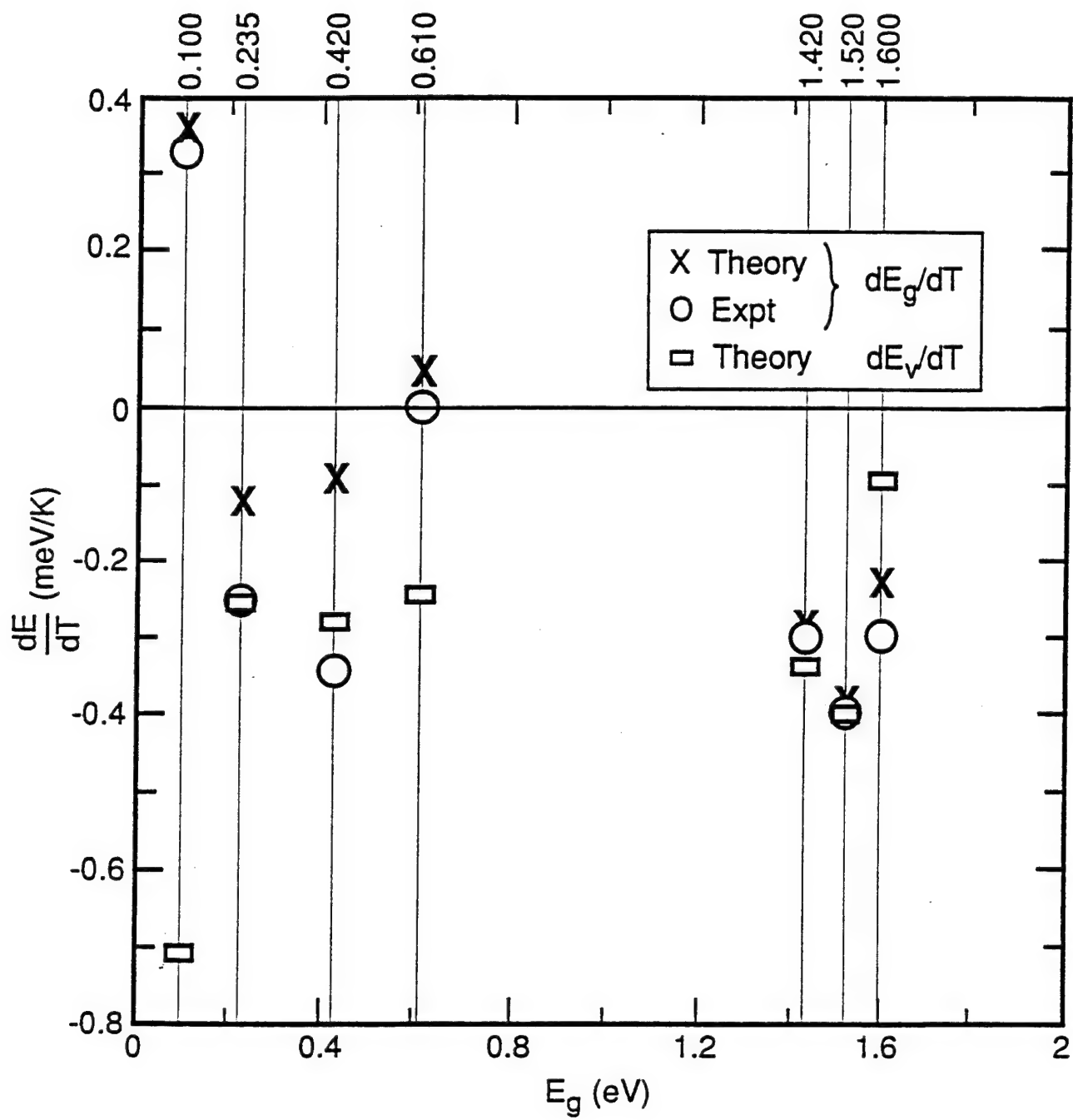
REFERENCES

1. Y.P. Varshini, *Phys.* 34, 149 (1967).
2. V. Heine and J.A. Van Vechten, *Phys. Rev. B* 13, 1622 (1976).
3. P.B. Allen and V. Heine, *J. Phys. C* 9, 2305 (1976).
4. P.B. Allen and M. Cardona, *Phys. Rev. B* 27, 4760 (1983).
5. S. Gopalan, P. Lautenschlager and M. Cardona, *Phys. Rev. B* 35, 5577 (1987).
6. M. Cardona and S. Gopalan, *Progress on Electron Properties of Solids*, eds. R. Girlanda et al. (Amsterdam, The Netherlands: Kluwer Academic Publishers, 1989), p. 52.
7. R.D. King-Smith, R.J. Needs, V. Heine and M.J. Hodgson, *Europhys. Lett.* 10, 569 (1989).
8. S. Zollner, S. Gopalan and M. Cardona, *Sol. State Comm.* 77, 485 (1991).
9. K.J. Malloy and J.A. Van Vechten, *J. Vac. Sci. Technol.* B 9, 2112 (1991).
10. P. Lautenschlager, M. Garriga, S. Logothetidis and M. Cardona, *Phys. Rev. B* 35, 9174 (1987), and references cited therein.
11. Z. Hang, D. Yan, F.H. Pollak, G.D. Pettit and M. Woodall, *Phys. Rev. B* 44, 10546 (1991).
12. L. Pavesi, F. Piazza, A. Rudra, J.F. Carlin and M. Ilegems, *Phys. Rev. B* 44, 9052 (1991).
13. E. Grilli, M. Guzzi, R. Zamboni and L. Pavesi, *Phys. Rev. B* 45, 1638 (1992).
14. P.Y. Liu and J.C. Maan, *Phys. Rev. B* 47, 16274 (1993).
15. M.E. Allali, C.B. Sorenson, E. Veje and P.T. Petersson, *Phys. Rev. B* 48, 4398 (1993).
16. G.L. Hansen, J.L. Schmit and T.N. Casselman, *J. Appl. Phys.* 53, 7099 (1982).
17. D.G. Seiler, J.R. Lowney, C.L. Littler and M.R. Loeoe, *J. Vac. Sci. Technol. A* 8, 1237 (1990).
18. J.C. Brice, *Properties of Mercury Cadmium Telluride*, eds. J. Brice and P. Capper (EMIS datareviews series 3, INSPEC publication, 1987), p. 105.
19. *Landolt-Bornstein Numerical Data and Functional Relationships in Science and Technology*, eds., Madelung, Schultz and Weiss, New series, Vol. 17 (1982).
20. A.-B. Chen and A. Sher, *Phys. Rev. B* 23, 5360 (1981).
21. S. Krishnamurthy, A.-B. Chen and A. Sher, *J. Appl. Phys.* 63, 4540 (1988).
22. S. Krishnamurthy, A. Sher and A.-B. Chen, *Phys. Rev. Lett.* 64, 2531 (1990); *Appl. Phys. Lett.* 55, 1002 (1989); *Appl. Phys. Lett.* 52, 468 (1988).
23. S. Krishnamurthy, A. Sher and A.-B. Chen, *Appl. Phys. Lett.* 53, 1853 (1988).
24. W. Harrison, *Electronic Structure and Properties of Solids* (San Francisco, CA: Freeman, 1980).
25. S. Krishnamurthy and M. Cardona, *J. Appl. Phys.* 74, 2117 (1993).
26. R.M. Martin, *Phys. Rev. B* 1, 4005 (1970).
27. S. Krishnamurthy and A. Sher, *J. Electron. Mater.* 24, 641 (1995).

Aug 6N







APPENDIX G

InTIP: A superior IRFPA alloy

A. Sher, M. van Schilfgaarde, A.-B. Chen, and S. Krishnamurthy

Proceedings of the IRIS Materials Meeting, August 1994

InTIP: A SUPERIOR IRFPA ALLOY

A. Sher, M. van Schilfgaarde, A.-B. Chen*, S. Krishnamurthy

SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025

*Auburn University
Auburn, Alabama 36849

ABSTRACT

$\text{In}_{1-x}\text{Tl}_x\text{P}$ is proposed as a promising material for infrared detectors. A number of key optical and structural properties are studied within local density-functional theory. $\text{In}_{1-x}\text{Tl}_x\text{P}$ at $x = 0.67$ and $\text{In}_{1-x}\text{Tl}_x\text{As}$ at $x = 0.85$ are estimated to have a gap of 0.1 eV. Their binding energies are larger than that of InSb, and they are found to form stable zincblende alloys for all x . $\text{In}_{1-x}\text{Tl}_x\text{P}$ nearly lattice matches to InP, and offers the potential to integrate detector array and read-out circuit.

We have proposed that the alloys $\text{In}_{1-x}\text{Tl}_x\text{P}$ and $\text{In}_{1-x}\text{Tl}_x\text{As}$ have properties that distinguish them as outstanding candidates for IR electro-optic receiver and emitter devices.¹ This paper concentrates on the properties of $\text{In}_{1-x}\text{Tl}_x\text{P}$ in the long wavelength infrared (LWIR) because it nearly lattice matches to InP substrates and, therefore, offers the prospect of integrated laser emitters, focal plane array (FPA) detectors, and read-out integrated circuits (ROIC) on the same chip. This capability could enable use of device architectures formerly deemed impractical because currently used LWIR materials are incapable of supporting them.

The properties of TIP, according to our first principals theory, that make it an attractive IR material candidate are:

- It forms in the zincblende structure.
- Its lattice constant (5.96 Å) closely matches that of InP (5.83 Å) (so the pseudo-binary $\text{In}_{1-x}\text{Tl}_x\text{P}$ liquidus and solidus phase diagrams have simple lens shapes).
- Its cohesive energy per atom (2.56 eV/atom) is 58% greater than that of HgTe (1.62 eV/atom).
- It is a semimetal with a negative gap of -0.27 eV, about the same as HgTe (-0.3 eV).

Table 1 presents the properties of the alloy with a 0.1 eV band gap that are related to LWIR-FPA performance and processing. The salient features are:

- The alloy concentration is $x = 0.67$, and the concentration variation of the gap $|dE_g/dx|$ is 1.42, 16% smaller than $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ (1.69).
- The elastic constants are ~33% larger than the LWIR HgCdTe alloy.
- The transverse optical phonon energy is 34.6 meV, 139% larger than HgCdTe (14.5), thereby limiting VLWIR utility to $\lambda_c < 36 \mu\text{m}$ (this is the only negative feature relative to HgCdTe).
- The temperature variation of the band gap² dE_g/dT near 77° K is small ($\sim -0.05 \text{ meV}/^\circ\text{K}$), about 1/7 as large as HgCdTe (0.36 meV/°K) (dE_g/dT for $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ vanishes near $x = 0.5$, while that of $\text{In}_{1-x}\text{Tl}_x\text{P}$ vanishes close to $x = 0.67$, the LWIR concentration, greatly simplifying designs for variable temperature operation and eliminating spatial variation in pixel performance caused by temperature gradients over array areas).
- The electron effective mass is 0.008, almost identical to HgCdTe (~ 0.008).
- The hole effective mass is 0.37, 43% smaller than HgCdTe (0.65) (which implies higher hole mobilities and substantially longer electron Auger recombination lifetimes for InTIP).
- The electron mobility at 80°K ($6 \times 10^4 \text{ cm}^2/\text{V-s}$) is 44% smaller than HgCdTe, but it does not die off as rapidly as temperature increases; consequently, electron mobility at 200°K is $4.5 \times 10^4 \text{ cm}^2/\text{V-s}$ while the same for HgCdTe is $2.24 \times 10^4 \text{ cm}^2/\text{V-s}$, only half as large. (This means the high temperature responsivity should not degrade as rapidly in InTIP.)

¹M. van Schilfgaard, A.-B. Chen, S. Krishnamurthy, and A. Sher, *Appl. Phys. Lett.*, in Press 1994.

²S. Krishnamurthy, A.-B. Chen, and A. Sher, submitted to *Appl. Phys. Lett.*, 1994.

This collection of properties--plus the extra ease of processing, the lower defect densities expected as a consequence of the high cohesive energy, and the superior InP substrate (three-inch diameter wafers with average dislocation densities $\sim 10^4 \text{cm}^{-2}$)³--lends support to the contention that $\text{In}_{1-x}\text{Tl}_x\text{P}$ will prove to be a striking LWIR-FPA material.

TABLE 1
LWIR MCT AND ITP PROPERTIES COMPARISONS

Property		Hg _{0.22} Cd _{0.78} Te		In _{0.33} Tl _{0.67} P	
		Theory	Experiment	Theory	Experiment
1	\bar{E}_g [eV]	0.1	0.1	0.1	—
2	\bar{E}_v [eV/atom]	1.66	1.75	2.75	—
3	\bar{a} [Å]	6.45	6.46	5.92	—
4	$\hbar\omega_{\text{To}}$ [meV]	14.5	14.12	34.6	—
5	B [10 ¹² erg/cm ³]	0.46	0.42	0.61	—
6a	dE_g/dx [eV] @ E_g (77°K) = 0.1 [eV]	1.71	1.69 @ 0°K	1.42	—
6b	dE_g/dT [meV/°K]	0.36	0.3	~-0.05	—
7	m_e^* @ 0° K	0.008	~0.009	0.008	—
8	m_h^* @ 0° K	0.65	0.38-0.71	0.37	—
9	μ_e [cm ² /V-s]				
9a	@ 80 K	1.07x10 ⁵	0.986x10 ⁵	6x10 ⁴	—
9b	@ 200 K	2.24x10 ⁴	2.0x10 ⁴	4.5x10 ⁴	—
10	μ_h [cm ² /V-S]				
10a	@ 77 K	—	600-1400	—	—
10b	@ 200 K	—	300-600	—	—
11a	τ_{eA} [ns] @ 10 ¹⁶ [cm ⁻³]	—	10	$\tau(\text{MCT}) < \tau(\text{ITP})$	—
11b	τ_{hA} [ns]	—	40	—	—

³D.F. Bliss, R.M. Hilton, and J.A. Adamski, *Journal of Crystal Growth*, **128**, 451, 1993.

ACKNOWLEDGMENTS

Partial support of this work from ONR Contracts N00014-88-C-0096 and N00014-89-K-0132 and ARPA Contract MDA972-92-C-0053 is gratefully acknowledged.

REFERENCES

- (1) M. van Schilfgaard, A.-B. Chen, S. Krishnamurthy, and A. Sher, *Appl. Phys. Lett.*, in Press 1994.
- (2) S. Krishnamurthy, A.-B. Chen, and A. Sher, submitted to *Appl. Phys. Lett.*, 1994.
- (3) D.F. Bliss, R.M. Hilton, and J.A. Adamski, *Journal of Crystal Growth*, 128, 451, 1993.

APPENDIX H

A study of the HgCdTe absorption edge at various temperatures

V. Ariel, V. Garber, G. Bahir, S. Krishnamurthy, and A. Sher

Abstract submitted to 1995 MCT Workshop

A Study of the HgCdTe Absorption Edge at Various Temperatures

V. Ariel, V. Garber, G. Bahir

Department of Electrical Engineering, Technion, Haifa 32000, Israel

S. Krishnamurthy*, A. Sher*

SRI International, Menlo Park, CA 94025

Optical properties of HgCdTe are extensively studied because it is the most important material for the detection of infra red radiation¹. In particular, the behavior of the absorption coefficient in the vicinity of the fundamental absorption edge is important because its spatial uniformity has a direct influence on HgCdTe focal plane array performance. In this work we present experimental and theoretical studies of the absorption coefficient in bulk and epitaxial HgCdTe layers at different temperatures.

Bulk samples used in this work were fabricated by Caminco and Technion. LPE grown samples were from Fermionics and MOCVD samples were from Soreq, NRC. The samples were cooled in the temperature range 20-300°K using a closed cycle liquid He refrigerator and investigated by FTIR transmission spectroscopy (Fig. 1). The first derivative of the absorption coefficient was used to study the band gap dependence on temperature in bulk samples² (Fig.2). The effect of temperature on the band gap grading in epitaxial layers was studied with the help of the second derivative (Fig. 3).

Absorption for photon energies $\hbar\omega$ less than the band gap E_g , which is known as the Urbach absorption tail³, is usually described as an exponential function⁴⁻⁷ of $\hbar\omega$. Our experimental data demonstrated a more complex dependence in this region of absorption which we were able to calculate by considering impurity tails of the conduction and valence bands⁸. The resulting semi-empirical expression, $\alpha = \alpha_0 \cdot \hbar\omega \times \exp[(\hbar\omega - E_g)/E_t]$, demonstrated agreement with experimental measurements in the Urbach range (Fig. 4) for all samples and in all temperature ranges.

In the fundamental absorption range, we considered the nonparabolicity of the conduction band which may be well represented⁹, up to about 0.5 eV from the conduction band edge, by $E_k = (\gamma k^2 + C^2)^{1/2} - C$, where C and γ are temperature and composition dependent parameters. This dispersion relation allows to calculate the absorption coefficient in the fundamental absorption range. Currently work is under way to compare this model with experimental measurements.

* Supported by ARPA/AFOSR contract MDA972-92-C-0053.

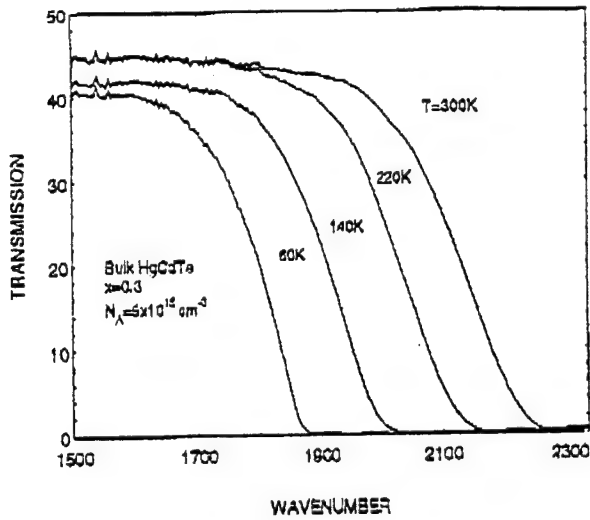


Fig.1

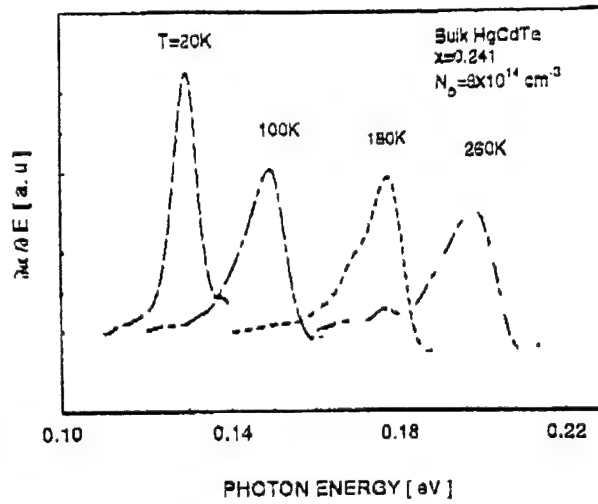


Fig.2

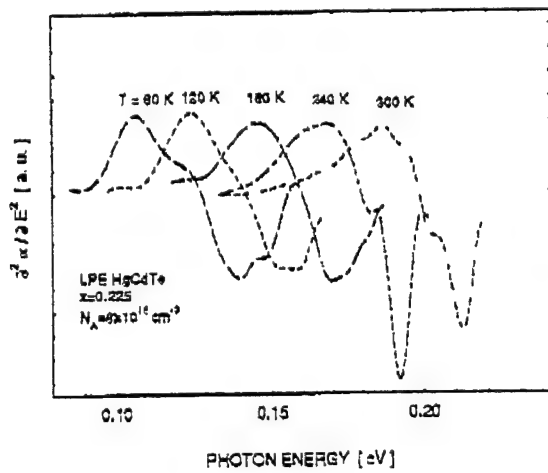


Fig.3

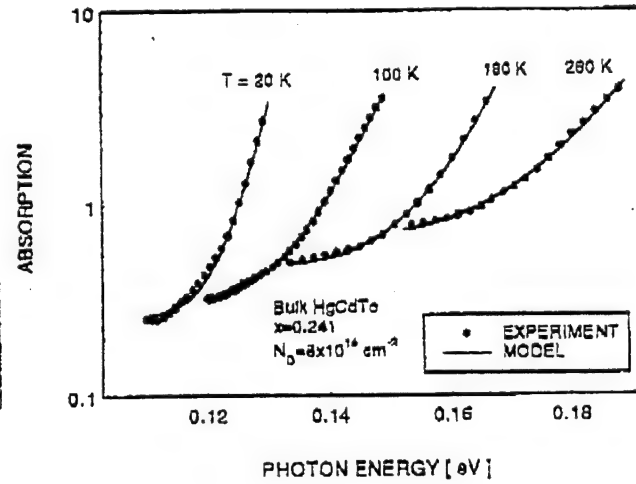


Fig.4

- 1 R. E. DeWames, J. M. Arias, L. J. Kozlowski, and G. M. Williams, *SPIE*, **1735**, 2, (1992).
- 2 V. Ariel, V. Garber, D. Rosenfeld, and G. Bahir, *Appl. Phys. Lett.*, **66**, 2101 (1995).
- 3 F. Urbach, *Phys Rev.* **92**, 1324 (1953).
- 4 E. Finkman, Y. Nemirovsky, *J. App. Phys.*, **50**, 4356 (1979).
- 5 C. A. Hougen, *J. Appl. Phys.*, **66**, 3763 (1989).
- 6 S. L. Price, P. R. Boyd, *Semicond. Sci. Technol.*, **8**, 842 (1993).
- 7 K. Liu, J. H. Chu, B. Li, D. Tang, *Appl. Phys. Lett.*, **64**, 2818 (1994).
- 8 V. Ariel, V. Garber, and G. Bahir, to be submitted to *Appl. Phys. Lett.*
- 9 S. Krishnamurthy, A. Sher, *J. Electr. Mater.* **24**, 643 (1995).

APPENDIX I

Transport studies in narrow-gap semiconductors revisited

S. Krishnamurthy and A. Sher

J. Electron. Mater. **24**, 641 (1995)

Transport Studies in Narrow-Gap Semiconductors Revisited

SRINIVASAN KRISHNAMURTHY and ARDEN SHER

SRI International, Menlo Park, CA 94025

Transport-related properties such as electron mobility, Hall coefficient, Fermi level, and energy gap are calculated with accurate analytical band structures, Fermi-Dirac statistics, and a full solution to the Boltzmann transport equation. These calculated values differ substantially from the ones obtained with parabolic or $k \cdot p$ generated band structure approximations for a $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ alloy. A new way to analyze absorption data to extract the temperature variation of the band gap is also explained.

Key words: Band gap, calculation of accurate band structures, electron mobility, Fermi level, Hall coefficient, HgCdTe

INTRODUCTION

Approximations such as parabolic band structures, Maxwell-Boltzmann (MB) statistics for electrons, and neglect of the gain term in the Boltzmann transport equation (BTE) (commonly known as collision time approximation) are often used to compare against and interpret experimental results. These approximations are often made to both elastic and inelastic scattering mechanisms. Even in large-gap materials, the constant effective mass approximation is valid only very near (within $\approx E_g/10$) to the band edge.^{1,2} This approximation is particularly poor for narrow gap materials, and nonparabolic corrections calculated in the $k \cdot p$ formalism are often used.³⁻⁵ Although this correction is substantial, it still differs considerably from our more accurately calculated band structures. However, our fit of these more accurate conduction bands to an analytical function makes many

results transparent and simplifies the calculations. As the Fermi energy can easily move into the conduction band of lightly doped small-gap materials, the form of the Boltzmann equation with Fermi-Dirac (FD) (instead of the usual MB) statistics must be used to obtain accurate transport coefficients.

In this paper, we report results from our study of absorption coefficients, Fermi energies, and Hall coefficients calculated with Fermi-Dirac statistics, accurate pseudopotential band structures fine tuned with tight-binding (TB) corrections, and the mobility with a full solution to the BTE.

BAND STRUCTURE

Quantitatively accurate band structures of most semiconductors⁶⁻⁸ can be obtained using a minimum set of sp^3 orbitals in semi-empirical calculations. First, for each alloy constituent, empirical pseudopotential form factors are used to calculate a TB Hamiltonian, H in the minimum set. This H is then transformed into a zeroth order Hamiltonian H_0 in an orthonormal

(Received October 13, 1993; revised August 15, 1994)

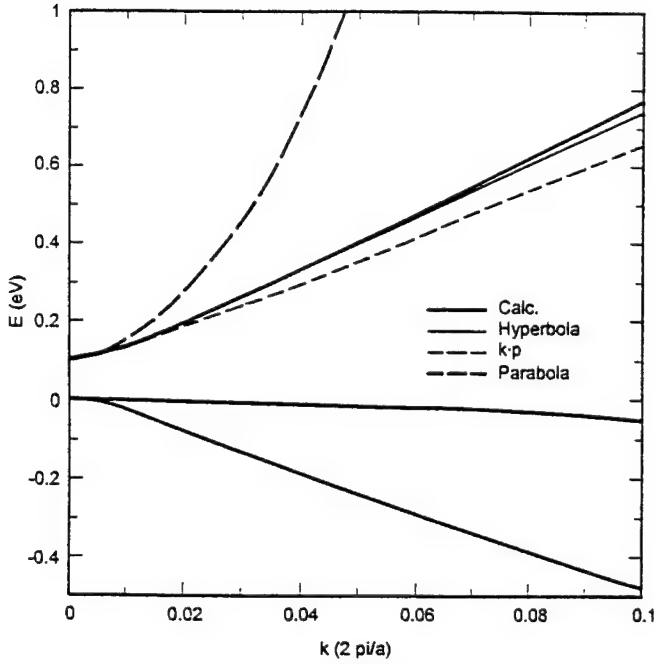


Fig. 1. Our approximation to 77K electronic structure of $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$.

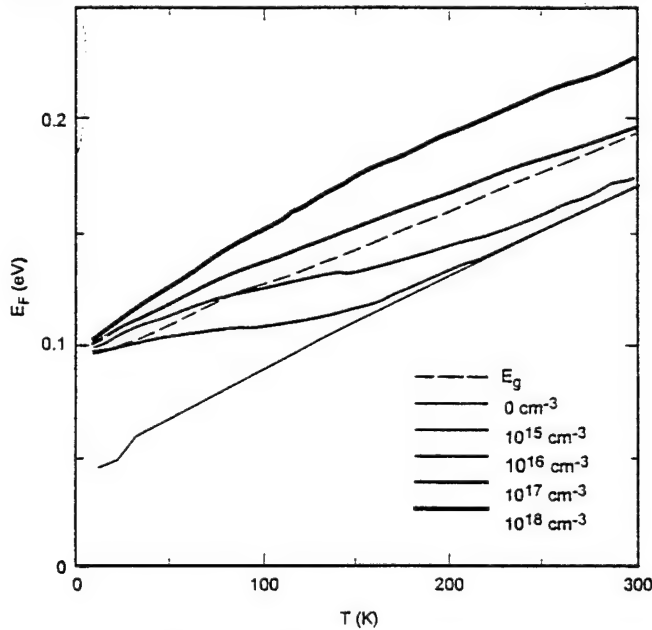


Fig. 2. Fermi energy as a function of temperature: hyperbolic bands. The dashed (experimental) curve is from Ref. 10.

basis. Then a perturbative Hamiltonian having a first-neighbor TB form is added to H_0 to fine tune the band structure to fit the measured symmetry point energies. Because long-range interactions are included in this Hamiltonian, the measured band curvatures are correctly reproduced. This procedure is followed for both HgTe and CdTe , and then the alloy band structures are calculated in the coherent potential approximation.

We focus on the $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ alloy with band-gap energy ranging around 100 meV at low temperature for the studies reported here. We find that the calculated conduction band is replicated very well by a

hyperbola,

$$E_k = (\gamma k^2 + c^2)^{1/2} - c \quad (1)$$

where γ and c are adjusted to fit the calculated band structure in the energy range of interest. When γ and c are treated as constants related to the band gap, E_g , and the effective mass, this expression reduces to the same nonparabolic correction form obtained in the $k \cdot p$ method.⁴ However, the numerical values of γ and c are not same as ours. For example, in the chosen case, γ and c are 48.3 and 0.058, respectively, whereas the corresponding $k \cdot p$ values are 36.0 and 0.05. The differences are found to be large enough to cause a noticeable change in the band structure and transport properties. The band structure calculated by diagonalizing the Hamiltonian is shown in Fig. 1 (thick line). We can see that the fitted hyperbola (thin line) agrees quite well up to an energy of 0.5 eV from the conduction band edge. Without loss of accuracy, in the studies considered here, Eq. (1) is used as the energy dispersion relation in transport expressions that follow. Also shown in Fig. 1 is the poor reproduction of the conduction band obtained with an effective mass approximation (long dashed line) and that with usual nonparabolic ($k \cdot p$) correction (short dashed line).

Two qualitative features of the band structure in Fig. 1 that impact transport properties should be noted. First, for energies $E - E_c$ greater than 50 meV where the shape of the conduction band is nearly linear in k the group velocity is nearly a constant independent of k . Then, the density of states (DOS) increases proportional to E rather than $E^{1/2}$ as in the case of parabolic bands. Clearly these features modify the transport properties of electrons occupying these states. As we will show in the following section, at the carrier concentration and temperatures often found in device structures, the Fermi level falls into the region where these features contribute to transport properties.

FERMI LEVEL

The calculation of the Fermi level, ϵ_F , as a function temperature, T and doping concentration, n_D is required for all transport calculations. A knowledge of temperature-dependent gap $E_g(T)$ is essential to obtain accurate values of ϵ_F in narrow gap material. Ideally, the temperature dependence should be deduced by including electron-phonon terms and lattice dilation in the Hamiltonian from which the variation of E_g with T can be obtained. We have developed a general method to incorporate the phonon and alloy effects into the same CPA formalism,⁹ but such an approach is not attempted here. Instead, we use the empirically deduced expression¹⁰ given by

$$E_g = 0.0954 + 0.327T/1000 \quad (2)$$

Although the gap has been fitted to a number of different analytical functions,¹⁰⁻¹² we chose the above expression simply to demonstrate the effects of various approximations. After obtaining trends, we in-

tend to repeat these calculations with a proper treatment by including the electron-phonon interaction Hamiltonian.

Here ϵ_F is calculated from the condition¹³ that at a given T the number of electrons in the conduction band is the sum of electrons excited from the valence band and the donor levels. In this study, where the modifications caused by the band structures are being emphasized, the donor states are assumed to be located at the bottom of the conduction band.¹⁴ The valence and conduction band DOS are calculated from our band structure. The valence band DOS yields an average hole effective mass of 0.65. The ϵ_F (relative to the valence band edge) as a function of T and n_D are given in Fig. 2. We found that the more accurate hyperbolic band is substantially different from parabolic-band generated values. The impact of these differences on the transport properties will be large when ϵ_F is located near the energy where the band changes its character from parabolic to linear. Also, when ϵ_F is near to or greater than E_g , the absorption cutoff wavelength is strongly influenced because the transition cannot take place to filled states. A proper account of this effect, called the Moss-Bernstein shift, must be included along with these accurate band structures. In addition, we note that the Fermi level and the measured band gap affect each other. The changes in the gap are accompanied by band curvature changes. The gap and the curvature change affect the Fermi energy, which in turn affects the apparent measured band gap. Hence, a proper interpretation of optical absorption to deduce the energy gap has to be done self-consistently to include both these effects.

ENERGY GAP

In order to emphasize the point that the measured value of E_g and its T dependence is sensitively dependent on the self-consistent determination of the band shapes and ϵ_F , we carried out a preliminary calculation of the absorption coefficient α . The absorption coefficient α is proportional to $k^2 \rho_c (1-f_c) \rho_v f_v$, where f is a k dependent FD distribution function, ρ is a k dependent DOS, and subscripts c, v represent conduction and valence band, respectively. The k^2 factor arises from the matrix elements that are in the expression for α . The temperature and energy-independent proportionality constant, which arises from the square of the overlap matrix element, is adjusted to agree with an experimental curve at 80K in the vicinity of a given value of $\alpha(\lambda)$ at wave length λ . A nearly linear dependence of $\log(\alpha)$ on λ was observed¹⁵ for various T between 80 and 300K and Cd concentrations near 0.22.

Because the band edges are often broadened by impurity and phonon scattering (known as Urbach tails), the gap cannot simply be assigned to the energy corresponding to the apparent cutoff wavelength λ_{co} where these curves project to zero. The procedure used by many authors¹¹ is to assign λ_{co} to be the place where α is 500 cm^{-1} or 1000 cm^{-1} . The justification for

this procedure is that the actual unbroadened shape of $\alpha(\lambda)$ is very sharp and if the Urbach tail ends at 500 cm^{-1} or 1000 cm^{-1} , then the corresponding λ at which this occurs will be close to the actual λ_{co} . In the spirit of this procedure, we adjust the proportionality constant to fit the $\log(\alpha)$ vs photon energy E curve at 80K. As seen from Fig. 3a, we find that this one constant fits nearly the entire curve at 80K. The constant is chosen such that the calculation and experiment agree very well in the vicinity of 500 cm^{-1} . There is little, if any, Urbach tail at this temperature. Once the constant is determined by this procedure, it is used for every temperature. At higher temperatures, the tails are present. We then artificially adjust the band gap

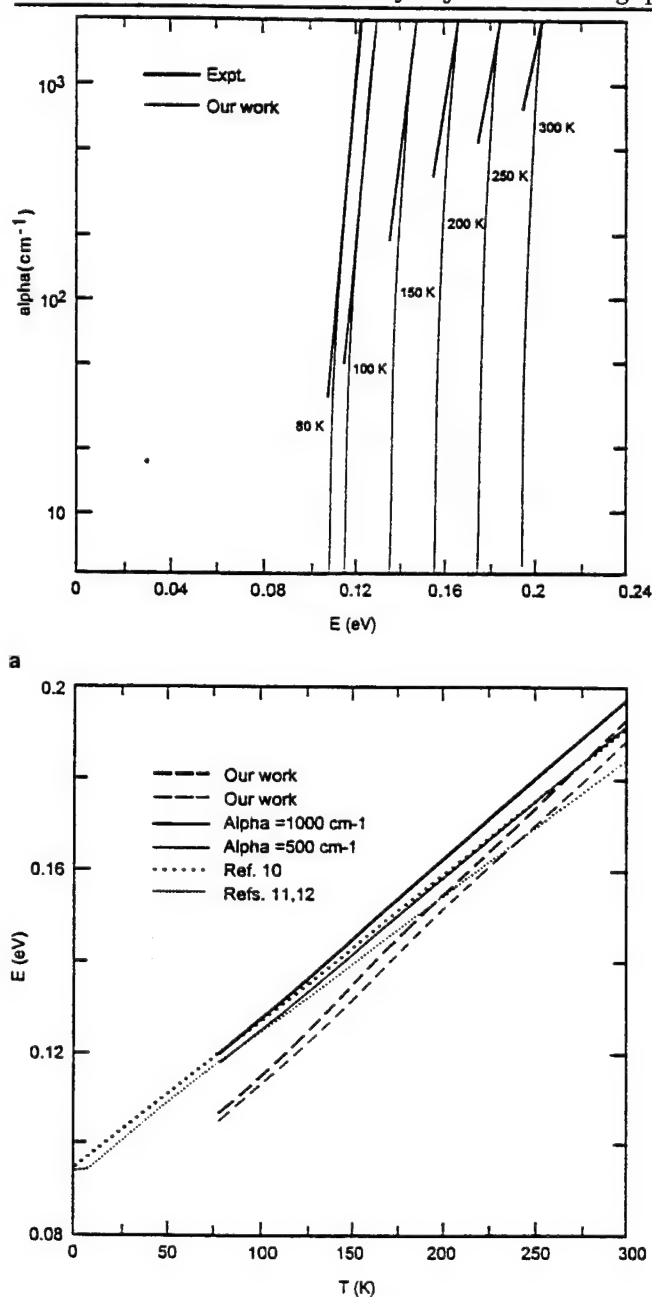


Fig. 3. (a) Absorption coefficient as function of photon energy; (b) energy gap as a function of temperature.

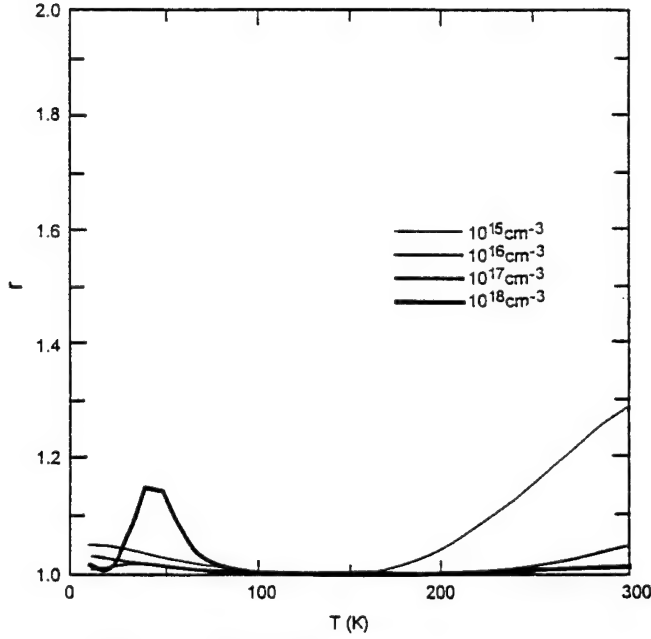


Fig. 4. Hall factor as a function of temperature: hyperbolic bands.

(with its corresponding band curvature) to mock the temperature variation by changing the Cd concentration until the calculated α curve fits a point on the experimental curve. In the course of this procedure, the Fermi energy also changes, which necessitates that the calculations be done iteratively until self-consistency is reached. The theoretical curve is then used to project the effective band gap. Fig. 3a displays the curves so obtained. The temperature variations of the energy gaps deduced in this way, with two different proportionality constants, are shown in Fig. 3b (dashed lines). Also shown in Fig. 3b (solid lines) for comparison are effective E_g vs T variations if values are taken from the data¹⁵ corresponding to α at 500 cm^{-1} and at 1000 cm^{-1} . Also, the experimental energy gap and its temperature variation measured by different groups¹⁰⁻¹² are plotted (dotted lines) in Fig. 3b for comparison. It is clear that various experiments differ substantially from each other. Also note that our curves are lower than the ones that are normally deduced from the same experiments, and the variation in T is no longer linear. Once the proportionality constant is calculated from our band structures, the energy gap is expected to lie between the two dashed lines shown here. However, the gap at $T = 0\text{K}$ predicted by our band structure method is in agreement with that of Refs. 10-12. The $E_g(T)$ will start with a zero slope¹² and connect smoothly to our 80K value. A nonlinear variation of E_g at much lower T , however, has been observed in recent two-phonon experiments.¹² All these results emphasize the need for proper calculations of $E_g(T)$, and $\alpha(T)$ so that more reliable device parameters can be established.

HALL COEFFICIENT

The carrier density, n in n -type material, is normally deduced from measurements of Hall coefficients R_H , given by r_e/en , by assuming the Hall factor

r_e is unity. If one uses a parabolic approximation for the conduction band and MB statistics, r_e is approximately unity. We set out to examine the effect of removing these approximations using the correct band structures. It requires generalizing the BTE to include FD distribution functions, $f(k)$. We start from

$$\frac{df(k)}{dt} = \sum_{k'} \left[w(k, k') f(k') (1 - f(k)) - w(k', k) f(k) (1 - f(k')) \right] \quad (3)$$

The first term of the right side is the gain term and the second one is the loss term. As a consequence of general statistical mechanics arguments,¹⁶ the ratio of transition probabilities is such that $w(k, k') e^{-\beta E_k} = w(k', k) e^{-\beta E_{k'}}$. In equilibrium, the left side of Eq. 3 is identically zero and f from Eq. 3 becomes the equilibrium FD distribution function f_0 given by

$$f_0(E_k) = \left(e^{\beta(E_k - \epsilon_F)} + 1 \right)^{-1} \quad (4)$$

where β is $(k_B T)^{-1}$. In the presence of electric and magnetic fields,

$$\frac{df(k)}{dt} = \frac{\partial f(k)}{\partial t} + \nabla f(k) \cdot \frac{e}{\hbar} (E + v \times B) \quad (5)$$

In steady state, the $\partial f(k)/\partial t$ in Eq. (5) vanishes. In the small field regime, we can linearize f and write it as a sum of f_0 and a perturbation $f_1(k)$. Discarding the derivative of $f_1(k)$ and after some algebraic manipulation, Eq. (5) reduces to¹⁷

$$\nabla f_0(k) \cdot \frac{e}{\hbar} (E + v \times B) = \sum_{k'} \left[W(k, k') f_1(k') - W(k', k) f_1(k) \right] \quad (6)$$

where the renormalized W and the usual transition probability per unit time w are related by

$$W(k, k') = w(k, k') \frac{(1 - f_0(k))}{(1 - f_0(k'))} \quad (7)$$

Note that for elastic scattering W and w are equal. However, for inelastic cases, the effect depends on whether the energies at k and k' are larger or smaller than ϵ_F . If both initial and final energies are larger (or smaller) than ϵ_F , then only a small correction to w is expected. However, if the initial state is above ϵ_F and the final state is below ϵ_F , then that scattering is suppressed.

In the collision time approximation, the gain term in Eq. (7) is neglected and the effective collision time, τ_k^F is

$$(\tau_k^F)^{-1} = \sum_{k'} W(k', k) \quad (8)$$

Using this collision time approximation, it is straightforward to obtain the expression for r_e .

$$r_e = 3k_B T \left[\frac{\sum_k f_0 \sum_k k^2 \gamma_k^3 (\tau_k^F)^2 f_0 (1-f_0)}{(\sum_k k^2 \gamma_k^2 \tau_k^F f_0 (1-f_0))^2} \right] \quad (9)$$

where $\nabla_k E_k$ is $\gamma_k k$. The r_e value calculated from Eq. (9) with parabolic bands is approximately 1 for higher temperatures but shows considerable structure with a maximum value of about 2 at lower temperatures. However, when the more accurate hyperbolic band structure is used, the variation at low T is reduced to a maximum value of about 1.2, suggesting that reported intrinsic densities may be smaller in this temperature range by approximately 20%. The calculated r_e with hyperbolic band structures are shown for various T and doping densities in Fig. 4. The effect of removing the collision time approximation and adding T dependence of band gap on the values of r_e still needs to be studied to extract correct carrier densities from Hall measurements. However, our tentative conclusion is that approximating r_e to be 1 is better than we had reason to expect.

DRIFT MOBILITY

The collision time approximation in the formalism developed above is removed, and a full solution to BTE with FD statistics¹⁷ is used to calculate the mobility μ . The details of this generalization are being published elsewhere and only the results are summarized here.¹⁷ This method is a generalization of the one to solve the Boltzmann equation with MB statistics.¹⁸

In Fig. 5, we compare the mobilities calculated in various approximations to experiments. The carrier concentration in the calculations was set equal to that used in the experimental value of $5 \times 10^{14} \text{ cm}^{-3}$. Two experimental data sets taken on LPE material^{19,20} are shown in Fig. 5 (dashed lines). The latest set exhibits higher mobilities for the same Cd content and carrier concentration and is presumably a better material. Also shown are the mobilities obtained from our hyperbolic band structure (thick solid line) and from the $k \cdot p$ band structure (thin solid line). For comparison, the mobility obtained in the collision time approximation with our band structure (dotted line) is also shown in Fig. 5.

It is instructive to compare various curves in Fig. 5. All curves are calculated with the same scattering parameters. Only ionized impurity and LO phonon scattering are included. No correction due to compensation is included. First, our calculated mobilities are higher than those from the $k \cdot p$ band structure. The smaller γ deduced from the $k \cdot p$ method means that the DOS is larger, resulting in this lower mobility. However, both curves predict a hump in the temperature variation of mobility near 40K, where phonon scattering takes over from the impurity scattering, which dominates at lower temperatures. The full solution to BTE, in conjunction with the change in the Debye

screening length and phonon scattering, gives rise to this hump. Second, the collision time approximation does not produce this hump. We note that the mobility calculated with a collision time approximation grossly overestimates the scattering rate and wipes out this peak in the mobility. A smaller peak near 200K is due to changes in the Fermi energy.

Our predictions fall within $\pm 25\%$ of the latest experimental values over the temperature range from 10 to 300K. Our calculated values are smaller at low T and larger at high T than experiments. As we demonstrated here, the electron mobility is a sensitive function of the shape of the band structure; we must await our better temperature-dependent band structures before improvements will be forthcoming. We have already shown from our preliminary studies of absorption coefficient that empirical gaps are as much as 20% too large. If detailed calculations verify these results, then impurity-dominated mobilities will increase at low T , and the small neglected scattering mechanisms (alloy disorder, transverse optical and acoustic phonons) will decrease the mobility slightly at high T to bring the predictions into better agreement with experiments at all temperatures.

CONCLUSIONS

In this paper, we have studied the effect of various approximations on electron transport coefficients and on ways to extract physical parameters from experiments. We point out how the values interpreted from experiments depend crucially on various approximations such as effective mass, MB statistics, and collision time. The main results are:

- Approximating the Hall factor by unity over a wide range of carrier concentrations and temperatures is accurate for most applications. An error of about 30% is expected at high T and low carrier concentration (10^{15} cm^{-3}), and about 20%

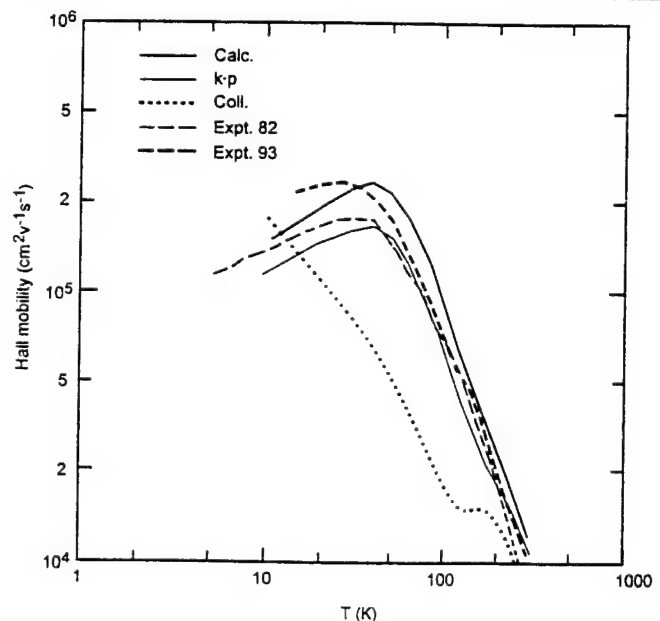


Fig. 5. Hall mobility as a function of T with various approximations.

is expected at low T and high carrier concentration (10^{18} cm^{-3}).

- The variation of the band gap with temperature for 22% Cd concentration is nonlinear and is faster at low temperature. The gap at low T is about 20 meV smaller than those usually quoted.
- The mobility calculated from a full solution to the BTE with FD statistics can explain the hump near 40K and is in good agreement with experiments. This result is obtained with no parameters in the theory adjusted to fit measured mobility data. Once the other scattering mechanisms such as acoustic and alloy disorder are included, the agreement is expected to be better.

ACKNOWLEDGMENT

We thank M.A. Berding and A.-B. Chen for many valuable discussions. The work was supported in part by ARPA contract MDA972-92-C-0053 and ONR contract N00014-93-0091.

REFERENCES

1. E.O. Kane, *J. Phys. Chem. Solids* 1, 249 (1957).
2. S. Krishnamurthy, A. Sher and A.-B. Chen, *J. Appl. Phys.* 61, 1475 (1987).
3. J.L. Schmidt, *J. Appl. Phys.* 41, 2876 (1970).
4. J.R. Meyer and F.J. Bartoli, *J. Vac. Sci. Technol.* 21, 237 (1982).
5. F.J. Bartoli, J.R. Meyer, R.E. Allen and C.A. Hoffman, *J. Vac. Sci. Technol.* 21, 241 (1982).
6. S. Krishnamurthy, A. Sher and A.-B. Chen, *Phys. Rev. B* 33, 1026 (1987).
7. A.-B. Chen and A. Sher, *Phys. Rev. B* 23, 5360 (1981).
8. M.A. Berding, S. Krishnamurthy, A. Sher and A.-B. Chen, *J. Vac. Sci. Technol.* A 5, 3014 (1987).
9. A.-B. Chen and A. Sher, *Phys. Rev. B* 5, 2897 (1972).
10. *Properties of HgCdTe*, J. Brice and P. Capper, eds., (New York: INSPEC, 1987) EMIS Data Review series No. 3, p. 103; and references cited therein.
11. G.L. Hansen, J.L. Schmidt and T.N. Casselman, *J. Appl. Phys.* 53, 7099 (1982).
12. D.G. Seiler, J.R. Lowney, C.L. Littler and M.R. Loloee, *J. Vac. Sci. Technol.* A 8, 1237 (1990).
13. S.M. Sze, *Physics of Semiconductor Devices* (New York: Wiley, 1981), p. 22.
14. See Ref. 17 for the case of donor level in the conduction band.
15. E. Finkman and Y. Nemirovsky, *J. Appl. Phys.* 50, 4356 (1979).
16. A. Sher and H. Primakoff, *Phys. Rev.* 119, 178 (1964).
17. S. Krishnamurthy and A. Sher, *J. Appl. Phys.* 75, 7904 (1994).
18. S. Krishnamurthy, A. Sher and A.-B. Chen, *Appl. Phys. Lett.* 55, 1003 (1989).
19. J. Bajaj, S.H. Shin, G. Bostrup and D.T. Cheung, *J. Vac. Sci. Technol.* 21, 244 (1982).
20. J. Bajaj, private communication (1993).

APPENDIX J

Transport studies in narrow-gap semiconductors revisited

S. Krishnamurthy and A. Sher

Proceedings of the IRIS Materials Meeting, August 1993

TRANSPORT STUDIES IN NARROW-GAP SEMICONDUCTORS REVISITED

August 1993

Srinivasan Krishnamurthy and Arden Sher
SRI International, Menlo Park, CA 94025

ABSTRACT

Transport-related properties such as electron mobility, Hall coefficient, Fermi level, and energy gap are calculated with accurate analytical band structures and Fermi-Dirac statistics. Calculated values differ substantially from the ones obtained with effective mass or parabolic band structure approximation of $\text{Hg}_{0.22}\text{Cd}_{0.78}\text{Te}$ alloy.

1. INTRODUCTION

The experimental results on electron transport properties of semiconductors are often compared to theory that rests on three approximations, namely, parabolic band structures for those states occupied in the measurement, Maxwell-Boltzmann (MB) statistics, and collision time approximation to a full Boltzmann gain-loss equation. These approximations are made to all scattering mechanisms, whether they are elastic or inelastic. It is well known that, even in large-gap materials, the constant effective mass approximation is valid only very near ($\approx E_g/10$) to the band edge.^{1,2} This approximation has been recognized to be particularly poor for narrow-gap materials and nonparabolic corrections calculated in the $\mathbf{k}\cdot\mathbf{p}$ formalism.³⁻⁵ This correction is substantial, but still differs considerably from our more accurately calculated band structures. In addition, our fit of the conduction band to an analytical function makes many results

transparent and simplifies the calculations. As the Fermi energy can easily move into the conduction band of lightly doped small-gap materials, the form of the Boltzmann equation with Fermi-Dirac (instead of the usual MB) statistics must be used to obtain accurate transport coefficients.

We report results from our study of absorption coefficient, Fermi energy, Hall coefficient, and electron mobility calculated with Fermi-Dirac statistics (FD), and an accurate pseudopotential band structure fine-tuned with tight-binding (TB) corrections.

2. BAND STRUCTURE

Quantitatively accurate band structures of group IV elements,⁶ III-V compounds,⁷ and II-VI compounds^{7,8} can be obtained using a minimum set of sp^3 orbitals in semiempirical calculations. First, empirical pseudopotential form factors are used to calculate a TB Hamiltonian, H in the minimum set. H is then transformed into a zeroth order H_0 in an orthonormal basis. Then, a perturbative Hamiltonian having a first-neighbor TB form is added to H_0 to fine-tune the band structure. Because long-range interactions are included in this Hamiltonian, the measured band curvatures are correctly reproduced. This procedure is followed for both HgTe and CdTe, and then the alloy band structures are calculated in the coherent potential approximation.

We focus $Hg_{0.22}Cd_{0.78}Te$ alloy with 100-meV band gap for the studies reported here. We find that the calculated conduction band is replicated very well by a hyperbola,

$$E_k = (\gamma k^2 + c^2)^{1/2} - c \quad (1)$$

where γ and c are adjusted to fit the calculated band structure in the energy range of interest. When γ and c are treated as constants related to the band gap, E_g and the effective mass, this expression reduces to the same nonparabolic correction form obtained in the $\mathbf{k} \cdot \mathbf{p}$ method.⁴ However, the numerical value of γ and c are not same as ours. For example, in the chosen case, γ and c are 48.3 and 0.058 respectively, whereas the corresponding $\mathbf{k} \cdot \mathbf{p}$ values are 36.0 and 0.05. The differences are found to be large enough to cause a noticeable change in the band structure.

The band structure calculated by diagonalizing the Hamiltonian is shown in Figure 1 (thick, solid lines). We can see that the fitted band structure (thin line) agrees quite well up to an energy of 0.5 eV from the conduction band edge. Without loss of accuracy, in the studies considered here, Eq. 1 is used as the energy-dispersion relation in the transport expressions. Also shown in Figure 1 is the poor reproduction of the conduction band obtained with an effective mass approximation (dashed line).

Two qualitative features of the band structure in Figure 1 that impact transport properties should be noted. First, for energies $E - E_c$ greater than 50 meV where the shape of the conduction band is nearly linear in k , the group velocity is a constant independent of the k . Then, the density of states (DOS) increases proportional to E rather than $E^{1/2}$ as in the case of parabolic bands. Clearly, these features modify the transport properties of electrons occupying these states. As we will show in the following section, at the carrier concentration and temperatures often found in device structures, the Fermi level falls into the region where these features contribute.

3. FERMIL LEVEL

The calculation of Fermi level ϵ_F as a function temperature T and doping concentration n_D is required for all transport calculations. A knowledge of temperature-dependent gap $E_g(T)$ is essential to obtain ϵ_F . Ideally, temperature dependence should be included in the Hamiltonian from which the variation of E_g with T could be obtained. We have developed a general method to incorporate phonon and alloy effects into the same CPA formalism,⁹ but such an approach is not attempted here. Instead, we use the expression¹⁰ given by

$$E_g = 0.0954 + 0.0327 T/1000 \quad (2)$$

After studying the effects of various approximations and obtaining trends, the calculations is repeated with the temperature-dependent Hamiltonian.

Then, the ϵ_F is calculated from the condition¹¹ that at a given T the number of electrons in the conduction band is the sum of electrons excited from the valence band and donor levels.

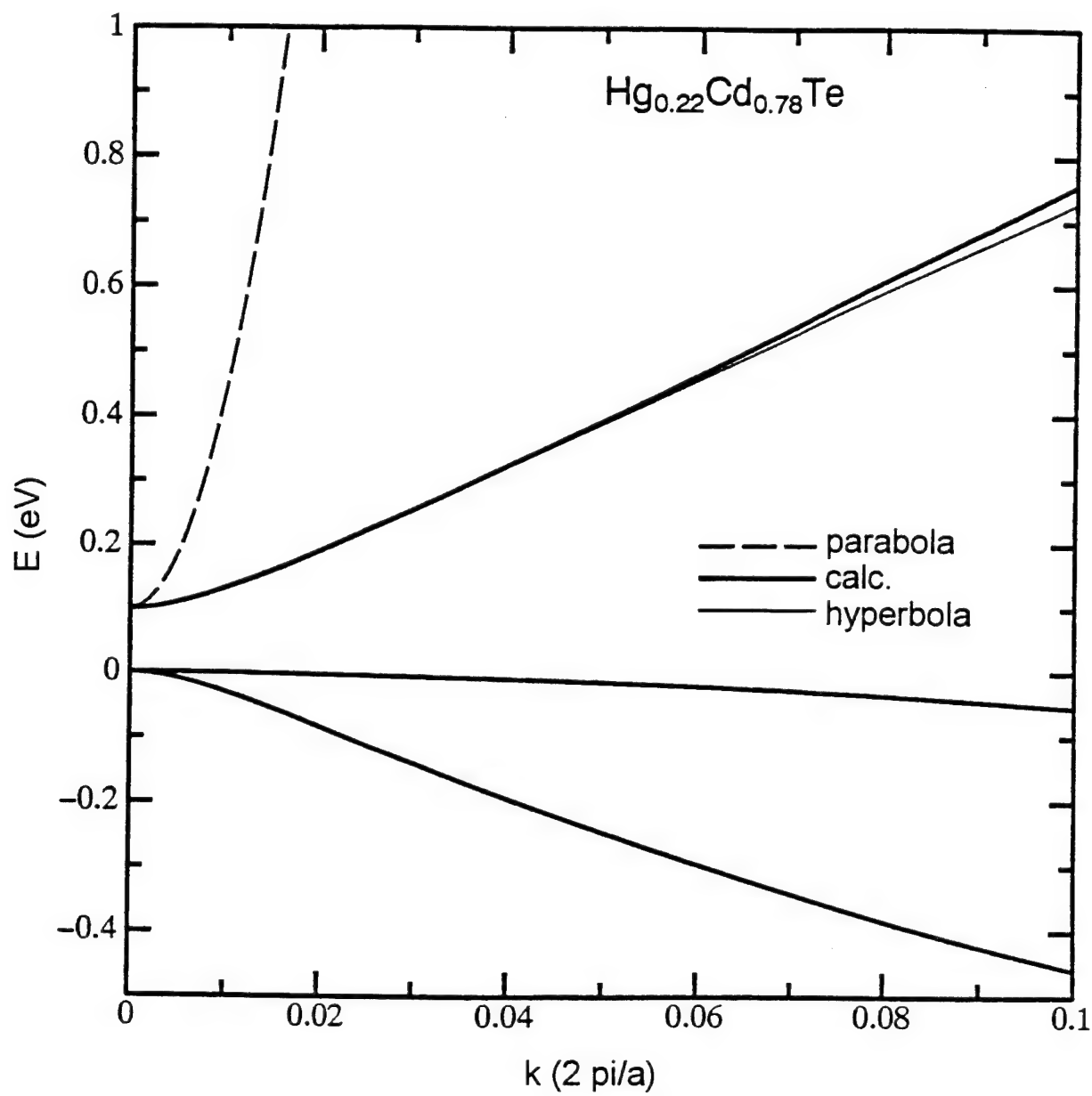


Figure 1. Electronic structure of HgCdTe

In this study where the modifications caused by the band structures are being emphasized, the donor states are assumed to be located at the bottom of the conduction band. The valence and conduction band DOS are calculated from our band structure. The valence band DOS yields a hole effective mass of 0.65. The ϵ_F (measured from valence band edge) as a function of T and n_D calculated from parabolic and hyperbolic band structures are given in Figure 2, a and b respectively. As expected, a hyperbolic band makes a substantial change from parabolic-band-generated values. The impact on the transport properties is large when ϵ_F is located at energy where the band changes its character from parabolic to linear. In addition, when ϵ_F is near to or greater than E_g , the absorption cutoff wavelength is strongly influenced. A proper account of the Moss-Bernstein shift must include these accurate band structures. We also note that the Fermi level and the measured band gap affect each other. Hence, a proper data reduction of cutoff wavelength determination of the energy gap has to be done self-consistently.

4. ENERGY GAP

To emphasize the point that the measured value of E_g and its T dependence is sensitively dependent on the knowledge of ϵ_F , we carried out a calculation of absorption coefficient α . The absorption coefficient α is proportional to $k^2 \rho_c (1 - f_c) \rho_v f_v$, where f is the FD distribution function, ρ is the DOS, and subscripts c, v represent conduction and valence band, respectively. The k^2 factor arises from the matrix elements that are in the expression for ϵ_F . The temperature- and energy-independent proportionality constant is adjusted to agree with the experimental curve at 80 K in the vicinity of a given value of α . Nearly linear dependence of $\ln(\alpha)$ on photon energy was observed¹² for T between 80 and 300 K and Hg concentrations near 0.22.

Because the band edges are broadened by impurity and phonon scattering (Urbach tail or broadening), the gap cannot simply be assigned to the energy wave length where these curves project to zero. The procedure used by many authors is to assign λ_{co} to be the place where α is 500 cm^{-1} or 1000 cm^{-1} . In the spirit of this procedure, we adjust our proportionality constant to fit the $\ln(\alpha)$ versus E curve at 80 K. As seen from Figure 3a, one constant fits the entire curve.

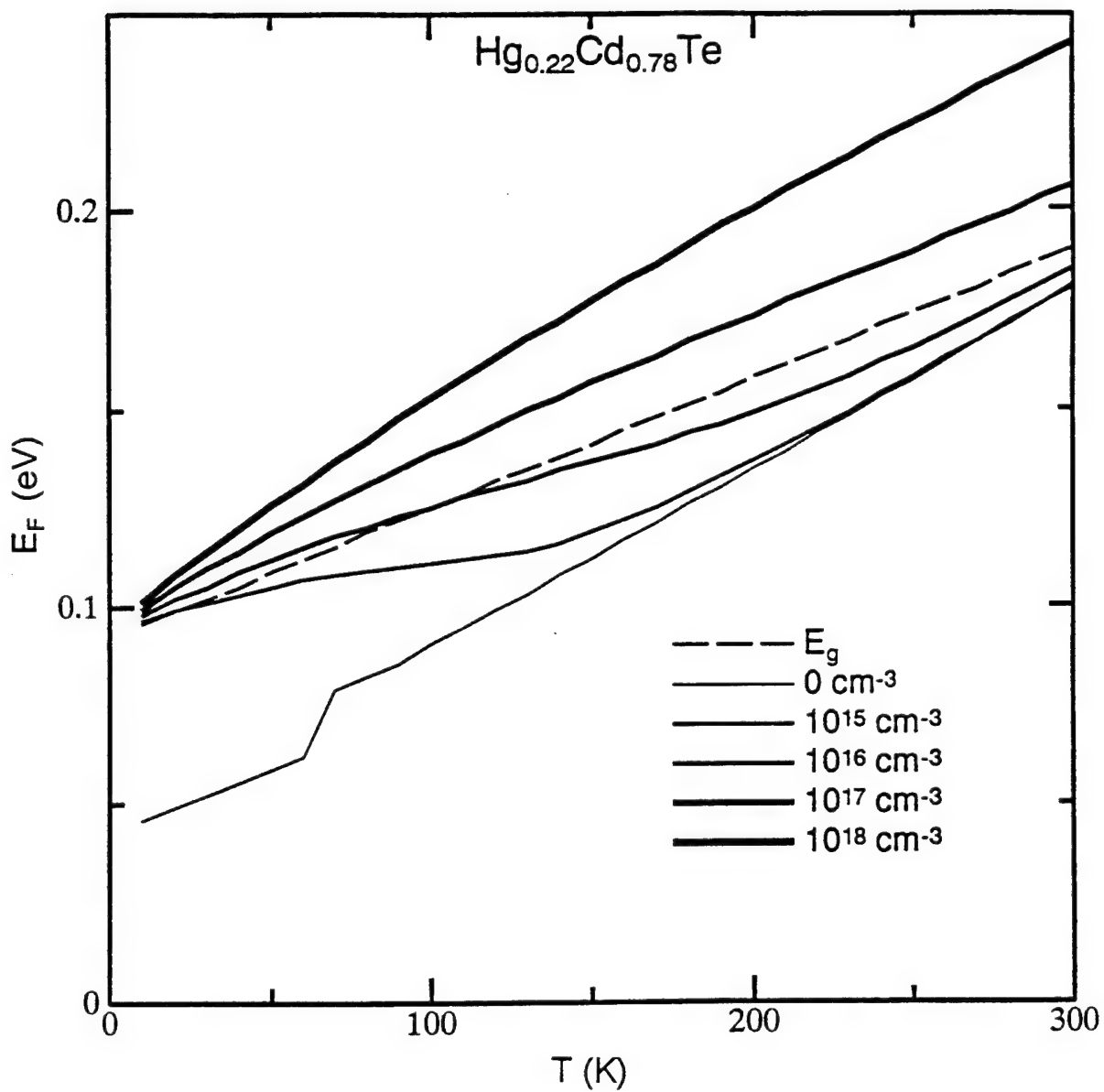


Figure 2a. Fermi energy as a function of temperature: parabolic bands

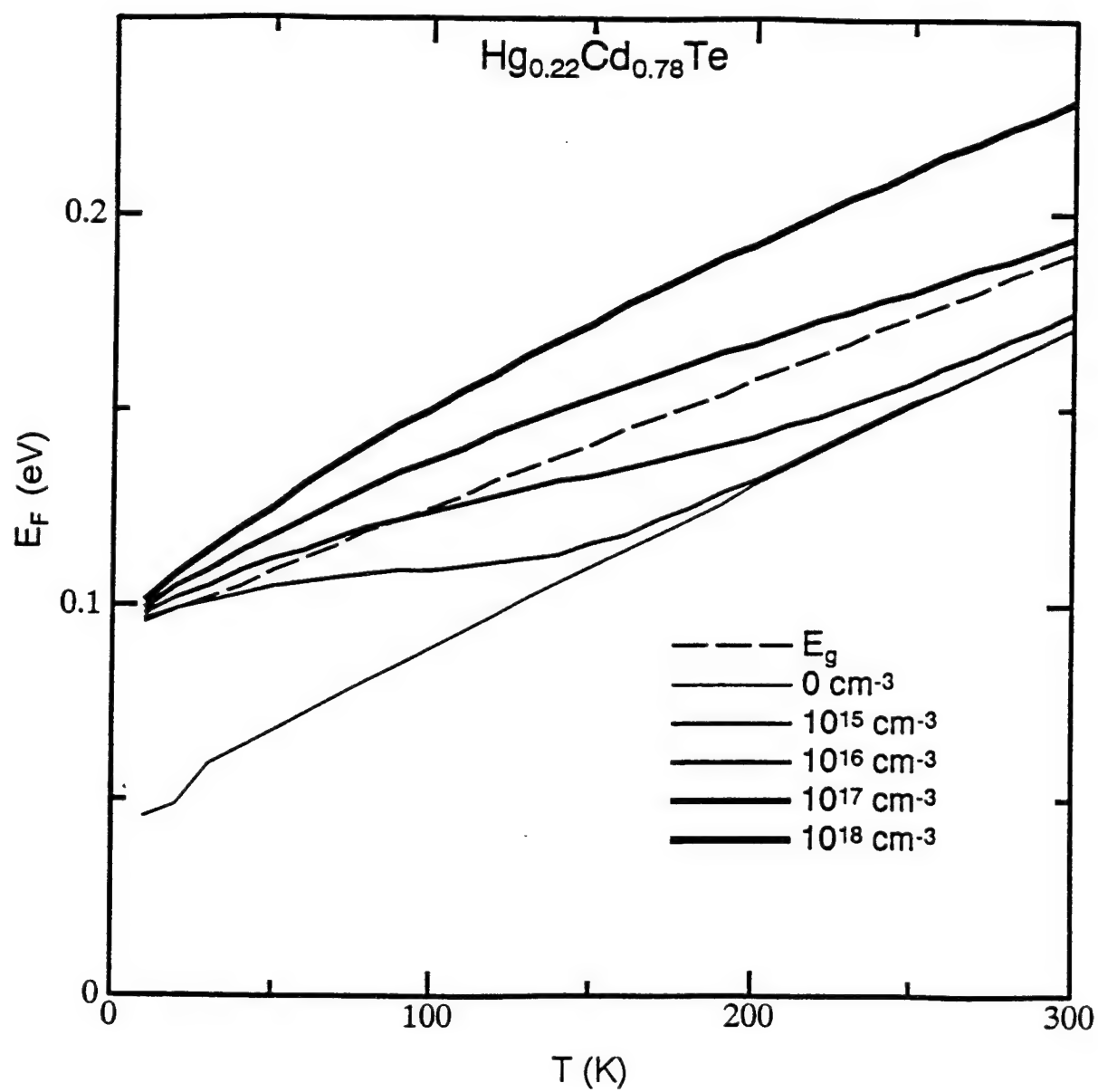


Figure 2b. Fermi energy as a function of temperature: hyperbolic bands

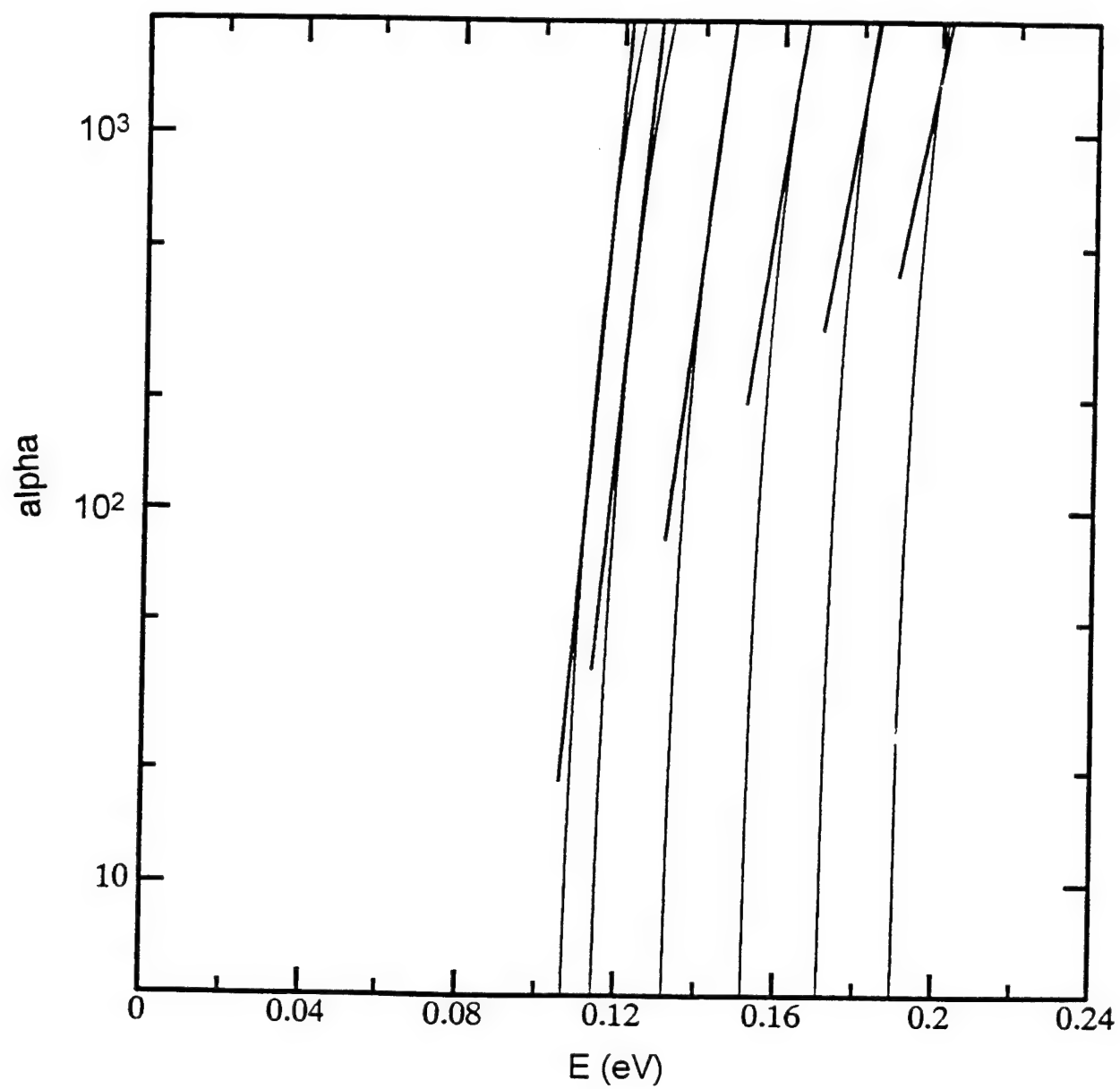


Figure 3a. Absorption coefficient as a function of photon energy

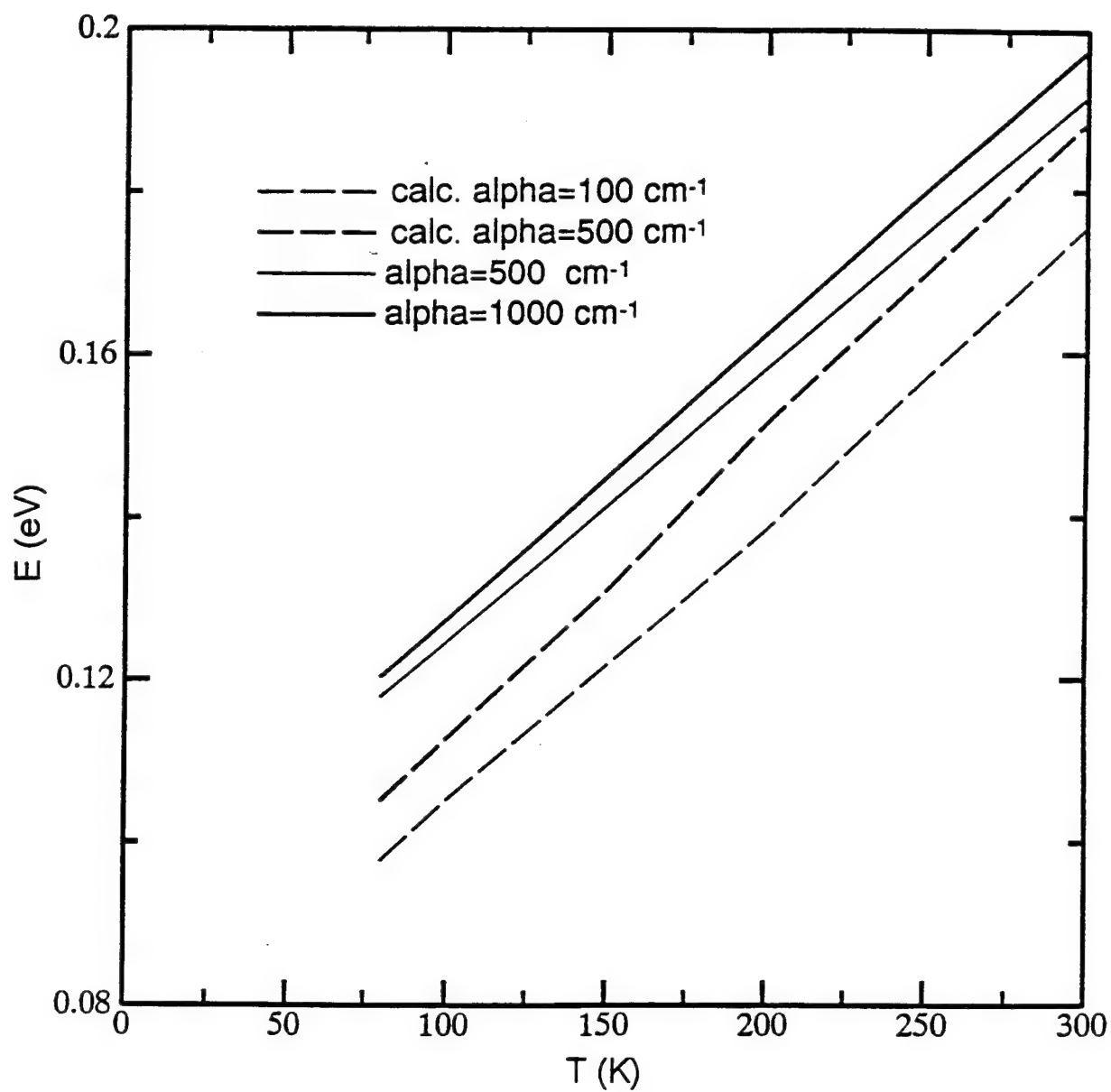


Figure 3b. Energy gap as a function of temperature

However, at higher temperatures the tail sets. We then recalculate the band structures and α for different band gaps (with corresponding curvature) until the curves at $\alpha = 500 \text{ cm}^{-1}$ for all temperatures, as shown in Figure 3a. The energy gap, variation with temperature, deduced by following this procedure is shown in Figure 3b (a dashed line). Also shown in Figure 3b (solid lines) for comparison are effective E_g versus T variations if values are taken from the data corresponding to $\alpha = 500 \text{ cm}^{-1}$ and 1000 cm^{-1} . Finally, because our procedure is ad hoc, we repeated our fits by adjusting to $\alpha = 100 \text{ cm}^{-1}$ values; the calculated energy gaps are also plotted in Figure 3b (a dashed line). We expect that the proper $E_g(T)$ curve lies somewhere between our two fit curves. Note that both curves are below the traditional curves and the variation in T is no longer linear, emphasizing the need for proper calculation of $E_g(T)$, $\alpha(T)$ so that more reliable device parameters can be established.

5. HALL COEFFICIENT

The carrier density, n in n-type material, is normally deduced from measurements of the Hall coefficient R_H , given by r_e/en , by assuming the Hall factor r_e is unity. If one uses parabolic approximation and MB statistics, r_e is approximately unity. We set out to examine the effect of removing these approximations, using the correct band structures. This requires generalizing the Boltzmann transport equation (BTE) to include the FD distribution function. We have

$$\frac{df(\mathbf{k})}{dt} = \sum_{\mathbf{k}'} [w(\mathbf{k}, \mathbf{k}')f(\mathbf{k}')(1-f(\mathbf{k})) - w(\mathbf{k}', \mathbf{k})f(\mathbf{k})(1-f(\mathbf{k}'))] \quad (3)$$

The first term of the right side of Eq. 3 is the gain term, and the second one is the loss term. In equilibrium, the left side of Eq. 3 is identically zero, and f becomes the equilibrium FD distribution function f_0 given by

$$f_0(E_k) = \left(e^{\beta(E_k - \epsilon_F)} + 1 \right)^{-1} \quad (4)$$

where β is $(k_B T)^{-1}$. In the presence of the electric and magnetic field

$$\frac{df(\mathbf{k})}{dt} = \frac{\partial f(\mathbf{k})}{\partial t} + \nabla f(\mathbf{k}) \cdot \frac{\mathbf{e}}{\hbar} (\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (5)$$

In steady state, the $\partial f(\mathbf{k})/\partial t$ in Eq. 5 vanishes. In the small-field regime, we can linearize f and write it as a sum of f_0 and a perturbation $f_1(\mathbf{k})$. Disregarding the derivative of $f_1(\mathbf{k})$ and after some algebraic manipulation, Eq. 5 reduces to

$$\nabla f_0(\mathbf{k}) \cdot \frac{\mathbf{e}}{\hbar} (\mathbf{E} + \mathbf{v} \times \mathbf{B}) = \sum_{\mathbf{k}'} [W(\mathbf{k}, \mathbf{k}') f_1(\mathbf{k}') - W(\mathbf{k}', \mathbf{k}) f_1(\mathbf{k})] \quad (6)$$

where the renormalized W and the usual transition probability per unit time w are related by

$$W(\mathbf{k}, \mathbf{k}') = w(\mathbf{k}, \mathbf{k}') \frac{(1-f_0(\mathbf{k}))}{(1-f_0(\mathbf{k}'))} \quad (7)$$

Note that for elastic scattering W and w are equal. However, for inelastic cases the effect depends on whether energies at \mathbf{k} and \mathbf{k}' are larger or smaller than ϵ_F . If both initial and final energies are larger (or smaller) than ϵ_F , only small correction to w is expected. However, if the initial state is above ϵ_F and the final state is below ϵ_F , that scattering is suppressed.

In the collision time approximation, the gain term in Eq. 7 is neglected and the effective collision time τ_k^F is

$$(\tau_k^F)^{-1} = \sum_{\mathbf{k}'} W(\mathbf{k}', \mathbf{k}) \quad (8)$$

Using this collision time approximation, it is straightforward to obtain the expression for r_e :

$$r_e = 3k_B T \left[\frac{\sum_{\mathbf{k}} f_0 \sum_{\mathbf{k}'} k^2 \gamma_k^3 (\tau_k^F)^2 f_0 (1-f_0)}{(\sum_{\mathbf{k}} k^2 \gamma_k^2 (\tau_k^F)^2 f_0 (1-f_0))^2} \right] \quad (9)$$

where $\nabla_{\mathbf{k}} E_{\mathbf{k}}$ is $\gamma_{\mathbf{k}} \mathbf{k}$. The calculated r_e with parabolic and hyperbolic band structures is shown for various T and doping densities in Figure 4, a and b respectively. In both cases, the r_e is approximately 1 for higher temperatures. At lower temperatures, it shows considerable structure with a maximum value of about 2 obtained in parabolic approximation. However, when more accurate hyperbolic band structure is used, the variation with T is reduced to a maximum value of about 1.2, suggesting that reported intrinsic densities may be smaller by approximately 20%. The effect of collision time approximation and T dependence of band gap on the values of r_e still

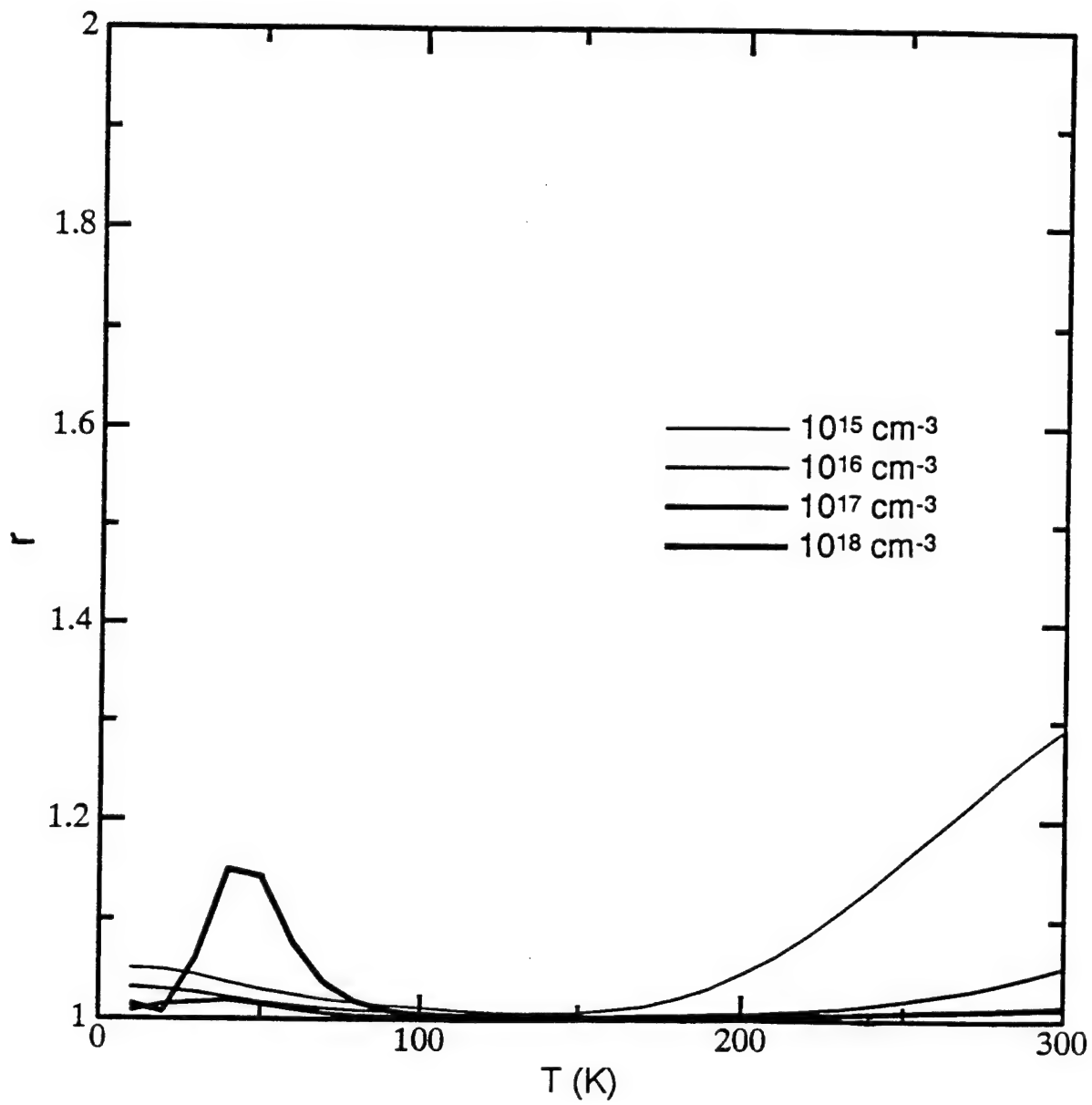


Figure 4a. Hall factor as a function of temperature: hyperbolic bands

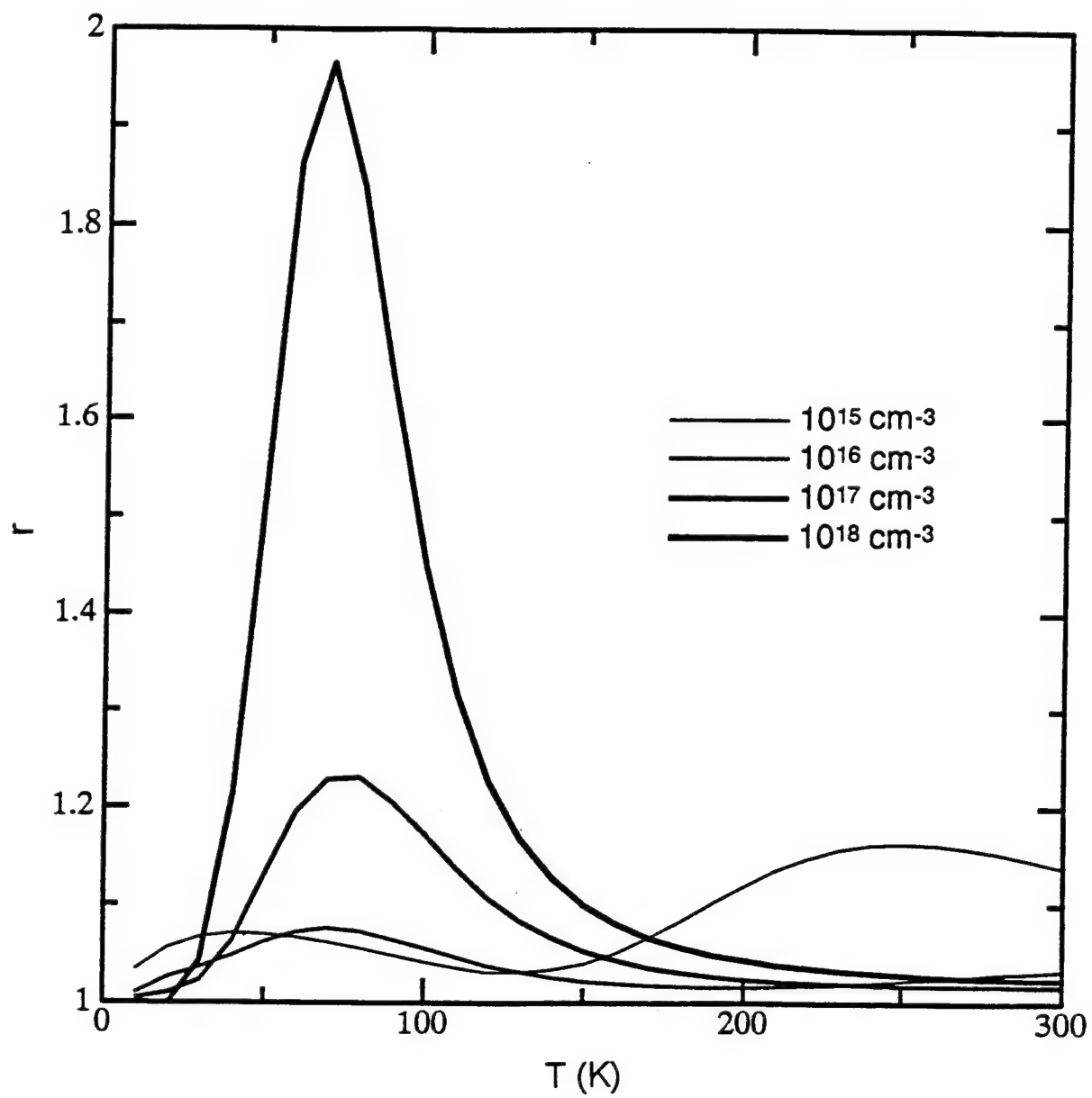


Figure 4b. Hall factor as a function of temperature: parabolic bands

need to be studied to extract correct carrier densities from Hall measurements. However, our tentative conclusion is that approximating r_e to be 1 is better than we had expected.

6. DRIFT MOBILITY

The formalism developed above to find a solution to BTE with FD statistics can be used to calculate the mobility, μ . We obtain the following expression in collision time approximation:

$$\mu = \frac{e}{3\hbar^2 k_B T} \left[\frac{\sum_{\mathbf{k}} k^2 \gamma_{\mathbf{k}}^2 \tau_{\mathbf{k}}^F f_0 (1 - f_0)}{\sum_{\mathbf{k}} f_0} \right] \quad (10)$$

The mobility calculated by including the scattering due to ionized impurities and polar optic phonons as a function of temperature and doping density is shown in Figure 5a. The change in the Debye screening length, complicated variation of Fermi level, and phonon scattering give rise to a crossover in mobility near 40 K, as seen in Figure 5a. However, it is well known that the collision time approximation always overestimates the velocity transition rate that determines the mobility.

For meaningful comparison with experiment, we generalized the above procedure to get full solution to BTE.¹³ Thus, calculated mobility with our hyperbolic band structure, and usual $\mathbf{k} \cdot \mathbf{p}$ band structure is shown along with that obtained in collision time approximation and compared with experiments¹⁴ in Figure 5b.

One important feature of Figure 5b is that the hump in mobility near 40 K could be explained with competing impurity and phonon scattering rates. Also note that collision time approximation grossly underestimates the mobility. Although $\mathbf{k} \cdot \mathbf{p}$ values are closer to experiment, the agreement is expected to be poor once the scattering mechanisms such as acoustic and alloy disorder and the effect of temperature-dependent band structures are included. These mechanisms lower the mobility calculated with hyperbolic band structures and yield an excellent agreement with experiment.

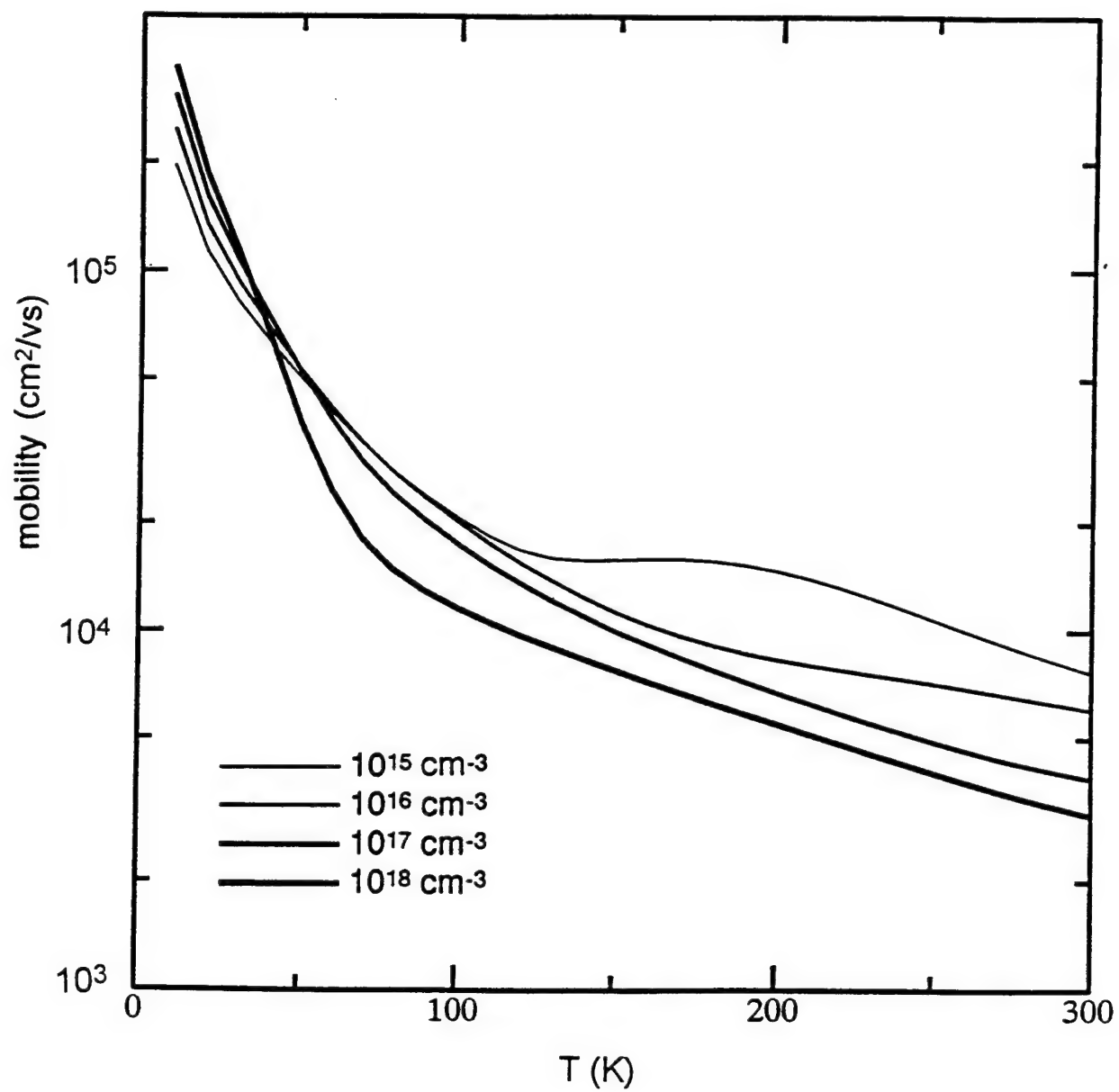


Figure 5a. Mobility with hyperbolic bands

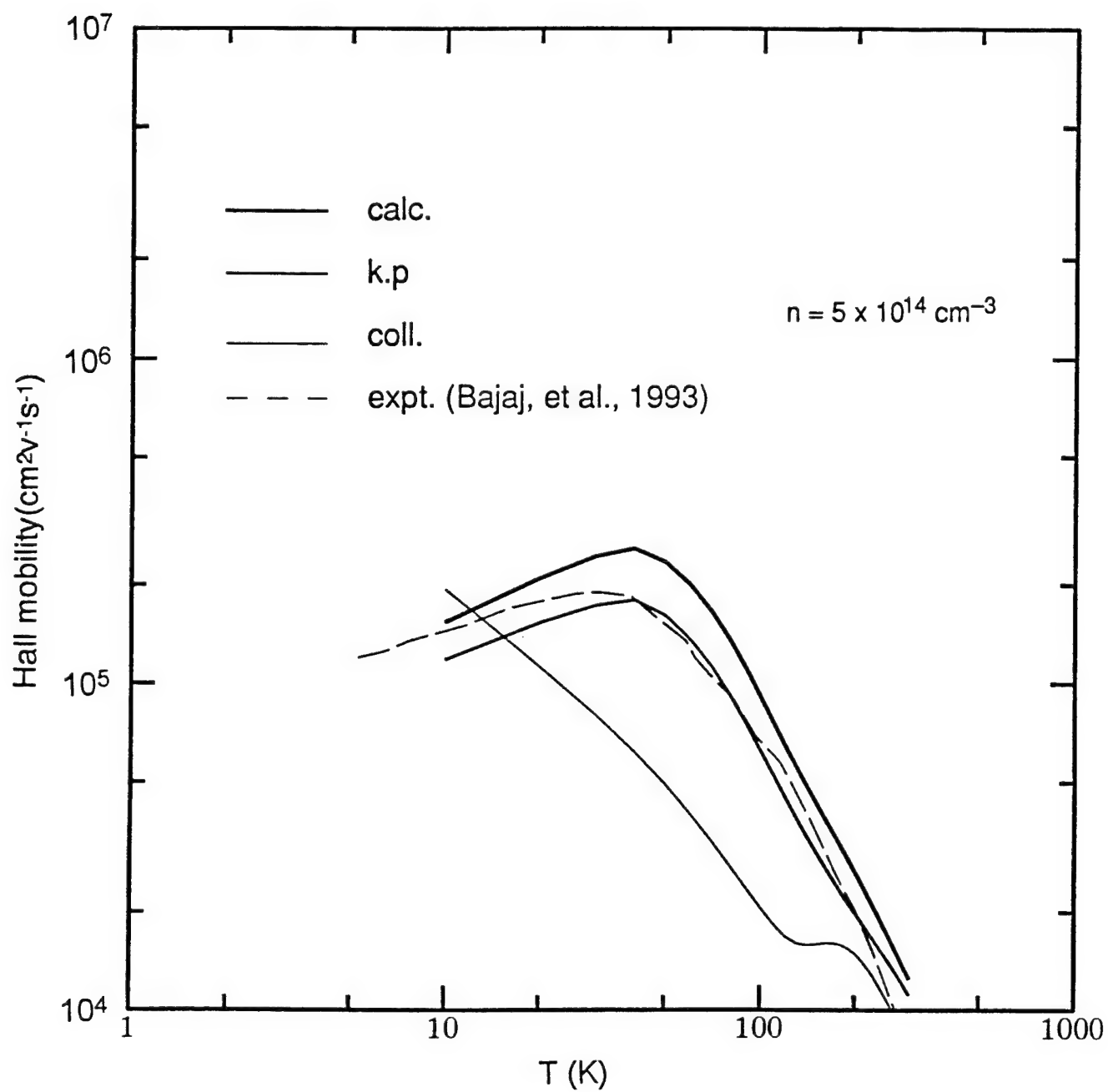


Figure 5b. Variation of Hall mobility with temperature

7. CONCLUSIONS

We have studied the effect of various approximations on electron transport coefficients and on ways to extract physical parameters from experiments. We point out how the values interpreted from experiments depend crucially on various approximations such as effective mass, MB statistics, and collision time. The results are

- Approximating the Hall factor by unity over a wide range of carrier concentrations and temperatures is accurate for most applications. An error of about 30% is expected at high temperatures and low carrier concentration (10^{14} cm^{-3}), and about 20% is expected at low T and high carrier concentration (10^{18} cm^{-3}).
- The variation of band gap with temperature is nonlinear and is faster at low temperature. The value at 22% Hg concentration is about 20 meV smaller than values usually quoted.
- The mobility calculated from full solution to BTE with FD statistics can explain the hump near 40 K and is in good agreement with experiments. Once the other scattering mechanisms such as acoustic and alloy disorder are included, the agreement is expected to be much better.

ACKNOWLEDGMENT

We thank Martha Berding for many valuable discussions and Dr. J. Bajaj for providing experimental mobility data. The work was supported in part by ARPA contract MDA972-92-C-0053 and ONR contract N00014-93-0091.

REFERENCES

1. Kane, E.O, 1957: J. Phys. Chem. Solids, **1**, 249.
2. Krishnamurthy, S., A. Sher, and A.-B. Chen, 1987: J. Appl. Phys., **61**, 1475.
3. Schmidt, J.L., 1970: J. Appl. Phys., **41**, 2876.
4. Meyer, J.R., and F.J. Bartoli, 1982: J. Vac. Sci. Technol., **21**, 237.
5. Bartoli, F.J., J.R. Meyer, R.E. Allen, and C.A. Hoffman, 1982: J. Vac. Sci. Technol., **21**, 241.
6. Krishnamurthy, S., A. Sher, and A.-B. Chen, 1986: Phys. Rev. B **33**, 1026.
7. Chen, A.-B., and A. Sher, 1981: Phys. Rev. B **23**, 5360.
8. Berding, M.A., S. Krishnamurthy, A. Sher, and A.-B. Chen, 1987: J. Vac. Sci. Technol., **A 5**, 3014.
9. Chen, A.-B., and A. Sher, 1972: Phys. Rev. B **5**, 2897.
10. Brice, J.C., 1986: *Properties of HgCdTe*, EMIS Datareview Series, No. 3, p. 103.
11. Sze, S.M., 1981: *Physics of semiconductor devices* (Wiley, New York), p. 22.
12. Finkman, E., and Y. Nemirovsky, 1979: J. Appl. Phys., **50**, 4356.
13. Krishnamurthy, S., and A. Sher 1993: The U.S Workshop on the Physics and Chemistry of HgCdTe and other IR materials, Seattle, Oct. 19-21.
14. Bajaj, J., S.H. Shin, G. Bostrup, and D.T. Cheung, 1982: J. Vac. Sci. Technol., **21**, 244; 1993, private communication.

APPENDIX K

Electron mobility in $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ alloy

S. Krishnamurthy and A. Sher

J. Appl. Phys. **75**, 7904 (1994)

Electron mobility in $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ alloy

Srinivasan Krishnamurthy and Arden Sher
SRI International, Menlo Park, California 94025

(Received 28 September 1993; accepted for publication 17 February 1994)

The electron mobility in $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ is calculated by solving the Boltzmann transport equation with Fermi-Dirac statistics and a full band structure. The calculated values are in excellent agreement with experiments, and effects of various traditional approximations are discussed.

I. INTRODUCTION

The experimental results on electron transport properties of semiconductors are often compared to theory that rests on three approximations, namely, parabolic band structures for those states occupied in the measurement, Maxwell-Boltzmann (MB) statistics, and the collision time approximation to full Boltzmann gain-loss equation solutions. These approximations are made to all scattering mechanisms whether they are elastic or inelastic. It is well known that, even in large-gap materials, the constant effective mass approximation is valid only very near (within $\sim E_g/10$) to the band edge.^{1,2} This approximation has been recognized to be particularly poor for narrow-gap materials, and nonparabolic corrections calculated in the $\mathbf{k}\cdot\mathbf{p}$ formalism are often used.³⁻⁵ This correction is substantial but still differs considerably from our more accurately calculated band structures. In addition, our fit of the conduction band to an analytical function makes many results transparent and simplifies the calculations. As the Fermi energy can easily move into the conduction band of lightly doped narrow-gap materials, the form of the Boltzmann equation with Fermi-Dirac (FD)—instead of the usual MB—statistics must be used to obtain accurate transport coefficients.

II. FORMALISM

In the literature, an iterative solution to the Boltzmann transport equation (BTE) with FD statistics has been derived.⁶ Although this method can yield nearly exact solutions by including more terms in the expansion and with more iterations, it can be time consuming since the procedure has to be repeated for every crystal wave vector \mathbf{k} . In this article, we derive an alternate method that first simplifies BTE with FD statistics to the point where the solution for full band structures can be obtained by successive substitution. This method is applied to the study of electron mobility as a function of temperature T and impurity concentration n_D in a $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ alloy. Results are compared with experiments.

We start from the BTE of a homogeneous medium obeying FD statistics. We have

$$\frac{df(\mathbf{k})}{dt} = \sum_{\mathbf{k}'} \{w(\mathbf{k}, \mathbf{k}')f(\mathbf{k}')[1-f(\mathbf{k})] - w(\mathbf{k}', \mathbf{k})f(\mathbf{k}) \times [1-f(\mathbf{k}')]\}. \quad (1)$$

The first term of the right-hand side of Eq. (1) is the gain term, and the second one is the loss term. In equilibrium, the

left-hand side is identically zero and because, in general for a system of interest interacting with a heat bath⁷

$$w(\mathbf{k}, \mathbf{k}')e^{-\beta E_{\mathbf{k}'}} = w(\mathbf{k}', \mathbf{k})e^{-\beta E_{\mathbf{k}}},$$

f becomes the equilibrium FD distribution function f_0 given by

$$f_0(E_{\mathbf{k}}) = (e^{\beta(E_{\mathbf{k}} - \epsilon_F)} + 1)^{-1}, \quad (2)$$

where β is $(k_B T)^{-1}$. In the presence of an electric field,

$$\frac{df(\mathbf{k})}{dt} = \frac{\partial f(\mathbf{k})}{\partial t} + \nabla f(\mathbf{k}) \cdot \frac{e}{\hbar} \mathbf{E}. \quad (3)$$

In steady state, the $\partial f(\mathbf{k})/\partial t$ in Eq. (3) vanishes. The solution $f(\mathbf{k})$ can always be written as a sum of $f_0(\mathbf{k})$ and some deviation $\Delta(\mathbf{k})$. That is,

$$f(\mathbf{k}) = f_0(\mathbf{k}) + \Delta(\mathbf{k}). \quad (4)$$

Note that $\sum_{\mathbf{k}} \Delta(\mathbf{k})$ is zero for a sample with ohmic contacts because the number of electrons should remain constant. Combining Eqs. (1)–(4), we obtain

$$\nabla f(\mathbf{k}) \cdot \frac{e}{\hbar} \mathbf{E} = \sum_{\mathbf{k}'} \{W(\mathbf{k}, \mathbf{k}')\Delta(\mathbf{k}') - W(\mathbf{k}', \mathbf{k})\Delta(\mathbf{k}) + [w(\mathbf{k}', \mathbf{k}) - w(\mathbf{k}, \mathbf{k}')]\Delta(\mathbf{k})\Delta(\mathbf{k}')\}, \quad (5)$$

where the renormalized W and the usual transition probability per unit time w are related by

$$W(\mathbf{k}, \mathbf{k}') = w(\mathbf{k}, \mathbf{k}') \frac{[1-f_0(\mathbf{k})]}{[1-f_0(\mathbf{k}')]} \quad (6)$$

Note that for elastic scattering W and w are equal; however, for inelastic cases the relative size of W to w depends on whether energies at \mathbf{k} and \mathbf{k}' are larger or smaller than ϵ_F . If both initial and final energies are larger (or smaller) than ϵ_F , then only small corrections to w occur; however, if the initial state is above ϵ_F and the final state is below ϵ_F , then for that scattering event W is suppressed. This tends, for example, to decrease the contribution of inelastic scattering events involving phonon emission. Also notice that if the perturbation is small, then the Δ^2 term can be neglected and Eq. (5) resembles the traditional MB steady-state BTE where W plays the role of w .

We further expand $\Delta(\mathbf{k})$ in power series of an indexing parameter λ , which we will eventually set to unity. Thus,

$$\Delta(\mathbf{k}) = \sum_{n=1}^{\infty} f_n(\mathbf{k})\lambda^n. \quad (7)$$

Substituting Eq. (7) in Eq. (5) and noting that $f(\mathbf{k})$ has one additional term, we obtain

$$\begin{aligned} \sum_{n=0}^{\infty} \nabla f_n(\mathbf{k}) \cdot \frac{e}{\hbar} E \lambda^{n+1} &= \sum_{\mathbf{k}', m=1}^{\infty} [W(\mathbf{k}, \mathbf{k}') f_m(\mathbf{k}') \\ &\quad - W(\mathbf{k}', \mathbf{k}) f_m(\mathbf{k})] \lambda^m \\ &\quad + \sum_{\mathbf{k}', m, m'=1}^{\infty} [w(\mathbf{k}', \mathbf{k}) - w(\mathbf{k}, \mathbf{k}')] \\ &\quad \times f_m(\mathbf{k}) f_{m'}(\mathbf{k}') \lambda^{m+m'}. \end{aligned} \quad (8)$$

By equating the coefficients of the same power of λ on both sides, we get a series of equations:

$$\nabla f_0(\mathbf{k}) \cdot \frac{e}{\hbar} E = \sum_{\mathbf{k}'} [W(\mathbf{k}, \mathbf{k}') f_1(\mathbf{k}') - W(\mathbf{k}', \mathbf{k}) f_1(\mathbf{k})], \quad (9a)$$

$$\begin{aligned} \nabla f_1(\mathbf{k}) \cdot \frac{e}{\hbar} E &= \sum_{\mathbf{k}'} \{W(\mathbf{k}, \mathbf{k}') f_2(\mathbf{k}') - W(\mathbf{k}', \mathbf{k}) f_2(\mathbf{k}) \\ &\quad + [w(\mathbf{k}', \mathbf{k}) - w(\mathbf{k}, \mathbf{k}')] f_1(\mathbf{k}) f_1(\mathbf{k}')\}, \end{aligned} \quad (9b)$$

and so on. Knowing f_0 [from Eq. (2)], Eq. (9a) is solved for f_1 and the solution is used [Eq. (9b)] for f_2 and so on. This procedure can be continued to the required precision. We emphasize that for low electric fields, only the lowest-order equation (9a) needs to be solved to obtain accurate answers. When hot-electron effects are addressed, iterative solutions should be used.

This procedure reduces the original integro-differential equation (5) to that of solving the following integral equation,

$$C(\mathbf{k}) = \sum_{\mathbf{k}'} [W(\mathbf{k}, \mathbf{k}') g(\mathbf{k}') - W(\mathbf{k}', \mathbf{k}) g(\mathbf{k})], \quad (10)$$

where $C(\mathbf{k})$ is a known function and $g(\mathbf{k})$ can be any of the set $\{f_n\}$. One of the fastest ways to solve for $g(\mathbf{k})$ is by expanding⁸ it in terms of an orthonormal basis set $\{\Phi(\mathbf{k})\}$. That is, take

$$g(\mathbf{k}) = \sum_n a_n \Phi_n(\mathbf{k}). \quad (11)$$

Substituting Eq. (11) in Eq. (10), multiplying both sides by $\Phi_m(\mathbf{k})$ and then summing over \mathbf{k} yields a matrix equation $\mathbf{C} = \mathbf{A}\mathbf{W}$, where \mathbf{A} is a row matrix of the required expansion coefficients. We see that \mathbf{A} can be easily obtained by multiplying the square matrix \mathbf{W}^{-1} by the row matrix \mathbf{C} . The size of the matrix \mathbf{W} is determined by the number of functions in the basis set. If the basis set is complete, the solution is exact; however, if the perturbation is small, one can truncate the basis set and obtain accurate solutions with only a few functions. For applied electric fields up to 7 kV/cm, eight basis functions were shown to be sufficient.⁹ Hence, in principle, following this procedure all f_n and the final distribution function can be obtained.

The formalism developed above to find solutions to BTE with FD statistics can be used to calculate the mobility μ from the following expression:

$$\mu = \frac{e}{\hbar} \frac{\sum_n \mathbf{k} a_n v(\mathbf{k}) \Phi_n(\mathbf{k})}{\sum_{\mathbf{k}} f_0(\mathbf{k})}, \quad (12)$$

where $v(\mathbf{k})$ is the group velocity of an electron in the state \mathbf{k} .

We apply this method to a study of the electron mobility in a $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ alloy with 0.1 eV (at 77 K) band gap. We use a high-quality band structure in the calculation. Quantitatively accurate band structures of group-IV elements,¹⁰ III-V compounds,¹¹ and II-VI compounds^{11,12} can be obtained using a minimum set of sp^3 orbitals in semiempirical calculations. First, empirical pseudopotential form factors are used to calculate a tight-binding Hamiltonian, H in the minimum set, but with interactions between atoms retained to all ranges. H is then transformed into a zeroth-order H_0 in an orthonormal basis. Then a perturbative Hamiltonian having a first-neighbor tight-binding form is added to H_0 to fine tune the band structure. Because long-range interactions are included in this Hamiltonian, the measured band curvatures as well as symmetry point energies are correctly reproduced. This procedure is followed for both HgTe and CdTe, and then the alloy band structures are calculated in the coherent potential approximation.

III. RESULTS

We find that the lower part of the calculated conduction band needed for the transport calculations is replicated very well by a hyperbola,

$$E_k = (\gamma k^2 + c^2)^{1/2} - c, \quad (13)$$

where γ and c are adjusted to fit the calculated band structure in the energy range of interest. When γ and c are treated as constants related to the band gap E_g and the effective mass, this expression reduces to the same nonparabolic correction form obtained in the $\mathbf{k} \cdot \mathbf{p}$ method;⁴ however, the numerical values of γ and c are not the same as ours. For example, in the chosen case, γ and c are 48.3 and 0.058, respectively, whereas the corresponding $\mathbf{k} \cdot \mathbf{p}$ values are 41.2 and 0.05. The differences are found to be large enough to cause a noticeable change in the band structure and transport coefficients. The band structure calculated by diagonalizing the Hamiltonian is shown in Fig. 1 (heavy line). We can see that the fitted conduction-band structure (thin line) agrees quite well up to an energy level of 0.5 eV from the conduction-band edge. In the studies presented here, Eq. (13) can be used as the energy dispersion relation without loss of accuracy, in the transport expressions that follow. Also shown in Fig. 1 is the poor reproduction of the conduction band obtained with parabolic approximation (heavy dashed line) and the usual nonparabolic correction (thin dashed line).

Two qualitative features of the band structure in Fig. 1 that impact transport properties should be noted. First, for energies $E - E_c$ greater than 50 meV where the shape of the conduction band is nearly linear in \mathbf{k} , the group velocity is a constant independent of the \mathbf{k} . Then, the density of states (DOS) increases proportional to E rather than $E^{1/2}$ as in the

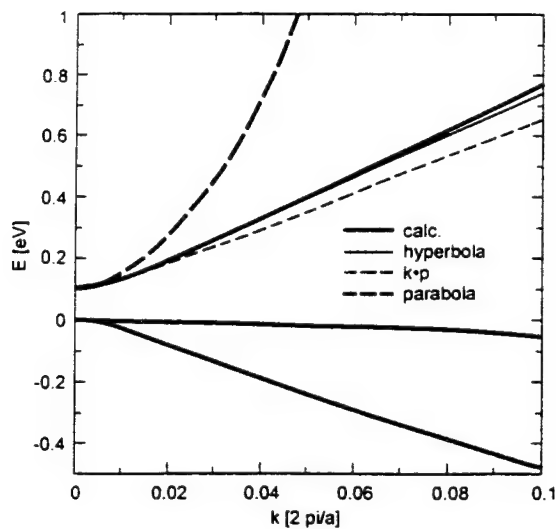


FIG. 1. Electronic structure of HgCdTe.

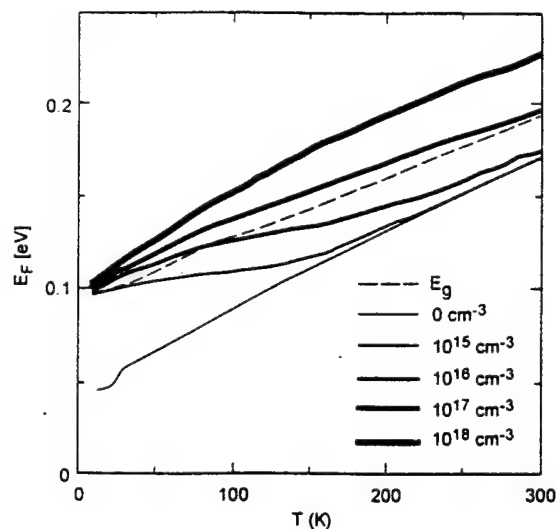
case of parabolic bands. Clearly, these features modify the transport properties of electrons occupying these states. The first feature, constant group velocity, eliminates increases in drift velocity arising from increases in electron temperatures. The second feature increases the scattering rates because the density of final states, into which the scattering can occur, is higher than that in a parabolic band. This effect decreases the drift velocity and mobility.

The calculation of the Fermi level ϵ_F as a function temperature T and doping concentration n_D is required for all transport calculations. A knowledge of temperature-dependent gap $E_g(T)$ is essential to obtain ϵ_F . Ideally, temperature dependence should be included by adding the electron phonon interaction to the Hamiltonian from which the variation of E_g with T is obtained. We have developed a general method to incorporate phonon and alloy effects into the same coherent potential approximation (CPA) formalism,¹³ but such an approach is not attempted here. Instead, we use an empirical expression first deduced by Hansen, Schmidt, and Casstleman¹⁴ given by

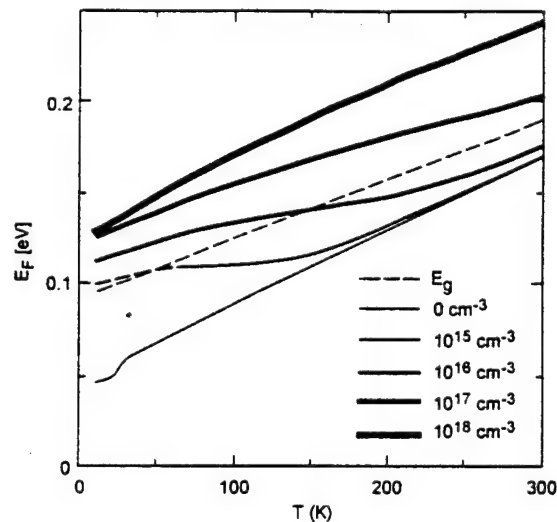
$$E_g = 0.0954 + 0.327T/1000. \quad (14)$$

After studying the effects of various approximations and obtaining trends, the calculations will be repeated with a proper temperature-dependent formalism.

Then ϵ_F is calculated from the condition¹⁵ that at a given T the number of electrons in the conduction band is the sum of electrons excited from the valence band and donor levels. The valence- and conduction-band DOS are calculated from our band structures. The valence-band DOS yields a hole effective mass of 0.65. We treat two cases, one where the donor states are assumed to be located at the bottom of the conduction band ($E_D=0$) so they can freeze out at low T , and a second where they resonate in the conduction band 30 meV above its bottom ($E_D=30$ meV) so the freeze-out is impossible until very high donor densities are reached. The ϵ_F (measured from the valence band edge) as a function of T and n_D , calculated from the hyperbolic band structures, is



(a)



(b)

FIG. 2. Fermi energy as a function of temperature: (a) $E_D=0$ eV; (b) $E_D=30$ meV.

given in Figs. 2(a) and 2(b), respectively, for $E_D=0$ and $E_D=30$ meV. The dashed line is the empirical band gap given by Eq. (2).

In the calculation of mobility we include the scattering due to ionized impurities and polar optic phonons whose scattering potentials are, respectively,

$$V_{\text{imp}} = \frac{e^2}{\epsilon_0 r} e^{-\lambda r},$$

$$V_{ep} = 4\pi i \left(\frac{e^2 \hbar}{2\xi_1 \omega_l V} \right)^{1/2} \sum_q a_q^* e^{-i\mathbf{q} \cdot \mathbf{r}} - a_q e^{i\mathbf{q} \cdot \mathbf{r}},$$

with

$$\lambda = \left(\frac{4\pi n_e e^2}{\epsilon k_B T} \right)^{1/2},$$

$$\xi_1^{-1} = \frac{\omega_l^2}{4\pi} (\epsilon_\infty^{-1} - \epsilon_0^{-1}),$$

where ϵ_0 , ϵ_∞ are zero and infinite frequency dielectric constants, $\hbar\omega_l$ is the longitudinal optical phonon energy, \mathbf{q} is the phonon wave vector, and a_q is the phonon annihilation operator. n_e is the number of electrons in the conduction band, which is, in general, larger than the number of ionized impurities. Note that the coupling constants are determined from fundamental properties of the material and we have no adjustable parameters to fit the transport data. Hermite polynomials have proved to be a good basis set for solving BTE with FD statistics. However, we find that only the first two basis functions are needed to obtain converged results. Hence, our basis set is e^{BE_k} and $\mathbf{k} \cdot \mathbf{E} e^{BE_k}$. The resulting mobility is plotted in Fig. 3 (heavy solid line).

Calculated mobilities for various impurity concentrations and temperature are shown in Figs. 3(a) and 3(b). In both cases, a peak in the mobility exists for a low concentration, in agreement with experiment.¹⁶ When the donor level is at the bottom of the conduction band, the Fermi energy lies well into the gap at low carrier concentration, and consequently only a few impurities are ionized. As the temperature is increased the number of electrons in the conduction band increases, and screening becomes more effective. The increase in the number of ionized impurities is overwhelmed by the increase in screening, and the impurity-limited mobility increases with T ; however, once the phonon scattering becomes stronger, the mobility decreases. This competition between impurity and phonon scattering gives rise to a peak in the mobility around 30 K. However, at still lower doping concentration, the impurity scattering is not effective at all, and the mobility then limited only by phonons will continue to increase as T is lowered. At higher doping densities the impurity-limited mobility also decreases with T because of a large increase in ionized impurities that now dominates the improved screening until very low temperatures are reached, and the peak vanishes. At intermediate doping densities a remnant of the peak is seen. At very high densities and very low temperatures, the screening once again wins, and along with the effects from the Fermi statistics the mobility increases slightly with n_D . This effect may be hard to verify experimentally as compensation will tend to mask it. Although a previous calculation¹⁷ suggests a stronger dependence of μ on n_D at low T , at least one experiment¹⁸ indicates a weaker dependence and the measured values are considerably scattered.

Such a trend did not reverse even when the donor levels are well in the conduction band. In addition, due to the absence of carrier freeze-out at low T , the impurity scattering is larger even at low concentration. The peak in mobility is obtained at $1 \times 10^{14} \text{ cm}^{-3}$, in Fig. 3(b). Although the impurity levels are resonant in conduction band, the line width may still be very small due to very small conduction-band DOS and, consequently, electrons still spend more time near impurities, giving rise to a screening larger than that considered

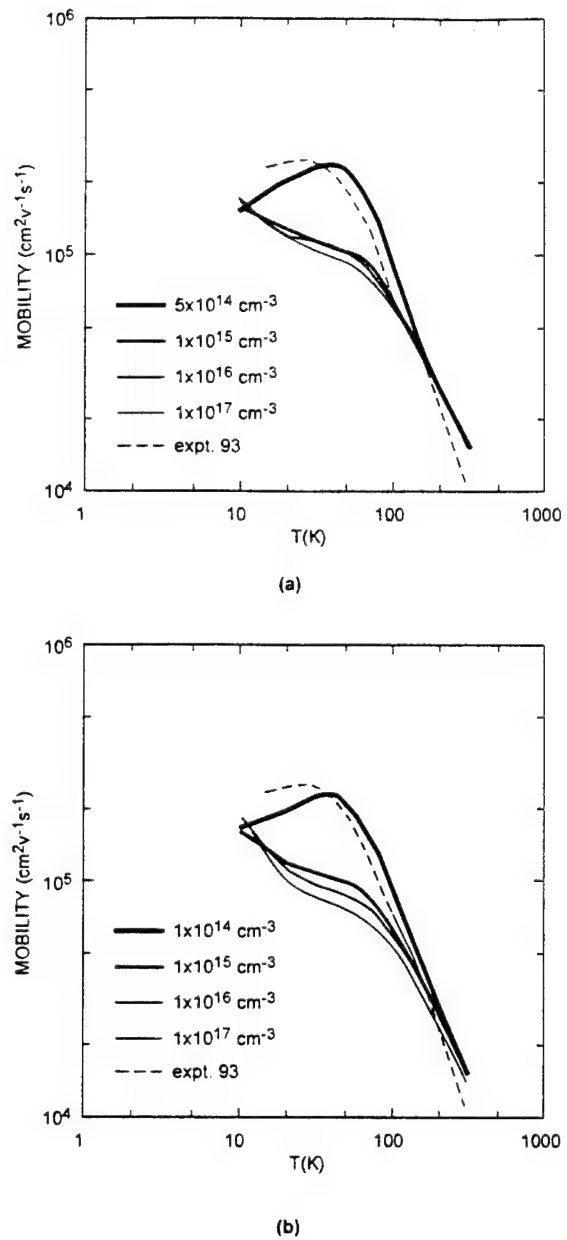
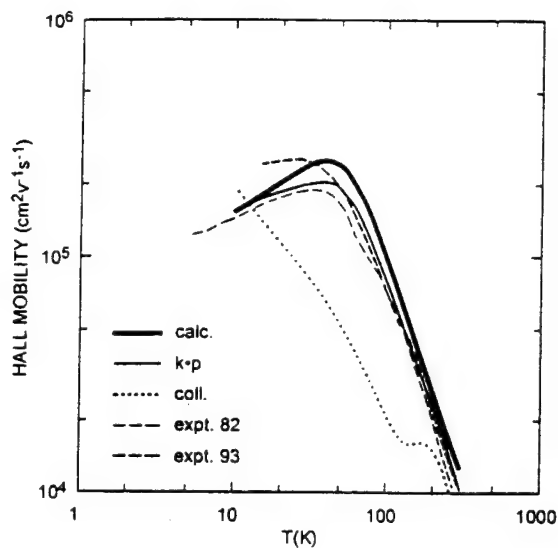


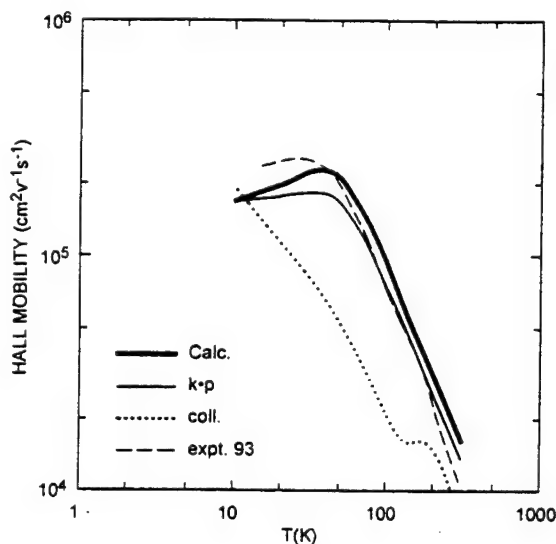
FIG. 3. Drift mobility as a function of T for various impurity concentrations: (a) $E_D = 0$ eV; (b) $E_D = 30$ meV.

in our calculation. If this sort of screening is considered, we will need a larger impurity concentration to obtain a peak in the mobility, yielding a better agreement with experiment. Such calculations are not attempted here. However, if the chosen impurity has donor levels near the conduction-band edge, the mobilities shown in Fig. 3(a) are appropriate.

In Fig. 4 we compare the mobilities calculated in various approximations to experiments. The carrier concentration in the calculations was set equal to that in experiments. Two experimental data sets taken on liquid-phase-epitaxy material^{16,19} are shown in Fig. 4 (dashed lines). The latest set exhibits higher mobilities for the same Hg content and carrier concentration and is presumably a better material. Also shown are the mobilities obtained from the hyperbolic band



(a)



(b)

FIG. 4. Hall mobility as a function of T with various approximations: (a) $E_D = 0$ eV; (b) $E_D = 30$ meV.

structure (heavy solid line) and from the $k \cdot p$ band structure (thin solid line). For comparison, mobility obtained in the collision time approximation (dotted line) is also shown.

It is instructive to compare various curves in Fig. 4. All curves are calculated with the same scattering parameters. First our calculated mobilities are higher than those from the $k \cdot p$ band structure. Smaller γ used in the $k \cdot p$ scheme means that DOS is larger, resulting in lower mobility. However, both curves predict a hump in the temperature variation of mobility near 40 K where phonon scattering takes over from impurity scattering, which dominates at lower temperatures. The full solution to BTE in conjunction with the change in the Debye screening length and phonon scattering give rise to this hump. Second, the collision time approximation does

not produce this hump. We note that the mobility calculated with a collision time approximation grossly overestimates the scattering rate and wipes out the peak in mobility. A smaller peak near 200 K is due to a change in the Fermi energy.

Calculated values of mobility are smaller at low T and larger at high T than in experiment. Since the curves in Fig. 4 clearly demonstrate that the electron mobility is a sensitive function in the shape of the band structure, we must await our better temperature-dependent band structures before improvements are forthcoming. We have already shown from our preliminary studies of absorption coefficient that empirical gaps are as much as 20% too large.²⁰ If detailed calculations verify these results, then impurity-dominated mobilities will increase at low T , and the neglected scattering mechanisms such as alloy disorder, transverse optical phonons, and acoustic phonons will decrease the mobility at high T to bring the predictions in better agreement with experiment.

IV. CONCLUSIONS

We have described another method to solve the Boltzmann transport equation with a full band structure and Fermi-Dirac statistics. This method is particularly useful when treating the transport properties of narrow-gap materials where the band structures are nonparabolic and Fermi statistics are essential. The method is applied to the study of electron mobility in a 100 meV band gap (at 77 K) $\text{Hg}_{0.78}\text{Cd}_{0.22}\text{Te}$ alloy. The calculated mobility compares well with the experimental mobility over a wide range of temperatures. Calculated values are 20% smaller at low T and 25% larger at high T than experiment. It appears that the calculated values will be smaller than experimental values at temperatures lower than 10 K. Most important, the observed hump in the mobility (with T) at low carrier concentration and the shape change at higher carrier concentration can be explained without resorting to additional scattering mechanisms or parameters. Further improvements await a proper calculation of the band structure and inclusion of alloy, transverse, and acoustic phonon scattering.

ACKNOWLEDGMENTS

We thank M. A. Berding and A.-B. Chen for many valuable discussions, and Dr. J. Bajaj for providing us with experimental mobility data. The work was supported in part by ARPA Contract No. MDA972-92-C-0053 and ONR Contract No. N00014-93-C-0091.

¹ E. O. Kane, *J. Phys. Chem. Solids* **1**, 249 (1957).

² S. Krishnamurthy, A. Sher, and A.-B. Chen, *J. Appl. Phys.* **61**, 1475 (1987).

³ J. L. Schmidt, *J. Appl. Phys.* **41**, 2876 (1970).

⁴ J. R. Meyer and F. J. Bartoli, *J. Vac. Sci. Technol.* **21**, 237 (1982).

⁵ F. J. Bartoli, J. R. Meyer, R. E. Allen, and C. A. Hoffman, *J. Vac. Sci. Technol.* **21**, 241 (1982).

⁶ D. L. Rode, *Semiconductors and Semimetals* (Academic, New York, 1972), p. 15.

⁷ A. Sher, and H. Primikoff, *Phys. Rev. B* **119**, 178 (1964).

⁸ M. Kohler, *Z. Phys.* **125**, 678 (1949).

⁹ S. Krishnamurthy, A. Sher, and A.-B. Chen, *Appl. Phys. Lett.* **55**, 1002 (1989); S. Krishnamurthy and M. van Schilfgaarde, *Computational Electronics* (Kluwer, Boston, 1990), p. 119.

¹⁰ S. Krishnamurthy, A. Sher, and A.-B. Chen, *Phys. Rev. B* **33**, 1026 (1986).

- ¹¹ A.-B. Chen and A. Sher, Phys. Rev. B **23**, 5360 (1981).
- ¹² M. A. Berding, S. Krishnamurthy, A. Sher, and A.-B. Chen, J. Vac. Sci. Technol. A **5**, 3014 (1987).
- ¹³ A.-B. Chen and A. Sher, Phys. Rev. B **5**, 2897 (1972).
- ¹⁴ J. C. Brice, Properties of HgCdTe, EMIS Data Review Series No. 3, 1986, p. 103.
- ¹⁵ S. M. Sze, *Physics of Semiconductor Devices* (Wiley, New York 1981), p. 22.
- ¹⁶ J. Bajaj (private communication).
- ¹⁷ J. J. Dubowski, T. Dietl, W. Szymanska, and R. R. Galazka, J. Phys. Chem. Solids **42**, 351 (1981).
- ¹⁸ W. Scott, J. Appl. Phys. **43**, 1055 (1972).
- ¹⁹ J. Bajaj, S. H. Shin, G. Bostrup, and D. T. Cheung, J. Vac. Sci. Technol. **21**, 244 (1982).
- ²⁰ S. Krishnamurthy and A. Sher, in Proceedings of IRIS Materials Specialty Group Meeting, Boston, 16 August, 1993.

APPENDIX L

Systematic study of gradient corrections in a wide range of elemental and compound materials

M. A. Berding and M. van Schilfgaarde

(In preparation for submission to *Phys. Rev. B*)

DRAFT

Systematic study of gradient corrections in solids

M. van Schilfgaarde, M. A. Berding and A. T. Paxton*

SRI International, Menlo Park, California 94025

**University of Belfast, Belfast, Ireland*

(September 6, 1995)

The total energy, lattice constant, bulk modulus, and various elastic constants for 58 elemental and compound solids are calculated using the local-density (LD) approximation and gradient corrections (GC) for exchange-correlation functional. For all systems, the GC functional of Lengreth and Mehl are compared with LDA calculations based on the von Barth-Hedin parameterization of the exchange-correlation functional. For a small group of materials, we have also used the Perdew-Wang GC functional, and compared the results to LDA calculations based on the Ceperly-Alder exchange-correlation functional. We find systematic improvement in the cohesive energies.

I. INTRODUCTION

Over the past few years there has been a renewed attention given the the generalized gradient corrections (GC) to the local density (LD) approximation. As its name implies, the LD is based on a local approximation for the exchange and correlation energies. The GC incorporated a non-local functional based on expansion of the exchange and correlation energy in terms of powers of the density gradient. Previous studies have focussed on a small set of materials, and no systematic study of a wide range of materials has been presented. That is our goal in this work.

Several forms of the GC have been proposed. In this paper we have focussed on the GC proposed by Langreth and Mehl¹. The nonlocal exchange and correlation functionals they developed were based on the random phase approximation (RPA) and therefore are used with the RPA values of the corresponding LD functionals².

We have systematically examined 58 elemental and compound semiconductors. The calculations done using the full-potential linearized muffin-tin orbital method³ in the local density functional approximation of Barth and Hedin.² In the FP-LMTO method, the only important approximation we make beyond the local-density approximation, lies in the treatment of the interstitial matrix elements. The LMTO method employs an atom-centered basis, represented by Hankel functions in the interstitial. For the calculations presented here, the basis consisted of a "triple kappa" basis 22 orbitals per atom, with energies -0.01 , -1 , and -2.3 Ry for the s and p orbitals, and -0.01 and -1 Ry for the d . Inside the muffin-tin (MT) spheres, wave functions are represented by spherical harmonics and numerically tabulated radial functions. The electron density and potential can be similarly represented, since the density generated by a Hamiltonian is obtained by summing over the eigenvectors. Outside the MT spheres, another treatment is necessary. Methfessel⁴ developed a simple, efficient way to represent the density and potential in the interstitial by extrapolation from the edges of MT spheres, where the value is well known. The electron density is represented in the interstitial as a linear combination of Hankel functions that are chosen to match the value and slope of the function at each MT sphere. Two Hankels per site and lm are enough to match the values and slopes at all MT spheres. This representation of the density throughout the interstitial is approximate, although it becomes exact near any MT sphere. Extensive tests show that the approximation works very well for close-packed systems, but the errors can become significant when the packing is poor. To ensure a good fit to the charge density and potential in the interstitial region of the zincblende solids, we include empty spheres at each tetrahedral interstitial site (rendering the sphere packing bcc for the ideal lattice). In addition, we added orbitals to the basis by centering them on the empty spheres. Addition of $2s$ and $2p$ orbitals changed the energy by approximately 0.1 mRy/atom, showing that the basis is nearly complete.

To assess the validity of the interstitial approximation for the representation of the charge density and interstitial matrix elements, an alternative approach was developed,³ which is similar to a procedure described by Jones.⁵ When calculated in this way the total energies changed by approximately 1 mRy/atom, showing that the approximation is a good one.

Both the charge density inside the spheres and the tails of Hankel functions centered on a neighboring sphere were expanded to $l = 6$. We estimate that the error introduced by truncation at $l = 6$ to be about 1 mRy/atom, in line with other errors in the method. The core was allowed to relax during the self-consistency cycle. The semicore d electrons in of the Group V and VI elements in the compound semiconductors were treated explicitly as valence states in a second panel; explicit treatment of these states was found to introduce a small but significant correction to the total energy. For the semiconductors, the Brillouin zone integrals were done by a sampling method for the charge

density, and the linear tetrahedron method for the band-structure energy, augmented by Blöchl weights, and a mesh of eight divisions was used (60 k-points). For the second panel six divisions were used corresponding to 28 k-points). Checks showed that this was sufficient to converge the energy to less than 0.1 mRy/cell.

We have bench marked our calculations with those of Filippi *et. al*⁶ done using a the highly accurate linear-augmented-plane wave-method.

II. RESULTS

In Fig. 1 we show a comparison of the atomic volume calculated using the LDA and GC. For nearly every system studied, the GC results predict larger atomic volumes. For systems like the Zn-bearing II-VI semiconductors and most of the 3d and early 4d and 5d transition elements for which LDA predicts atomic volumes that are substantially too small, the GC improve the agreement with experiment.

In Fig. 2 we have plotted the cohesive energy per atom, and have compared the results using the LDA and GC. For the semiconductors, the well known over binding of LDA of the order of 0.5 eV, is dramatically reduced using the GC. A similar reduction in the over-binding comparing GC to LDA is observed for the transition metals. The nearly abrupt shift in the errors in Fig. 2-b around the middle of the transition element series is most likely due to errors resulting from the free-atom spin polarization energy, which is quite big for these elements.

The experimental total energies reported in the top panels of Fig. 2 include contributions that are not included in the LDA and GC, in particular the zero point energy. This contribution is especially important for material with low atomic weight and large elastic constants. We have calculated the zero point energy using a valence force field model previously developed for calculation the vibration free energies of native point defects in semiconductors. In Table I we summarize the corrections for the zinc-blende semiconductors.

All of the solid cohesive energies reported in Fig. 2 are reference to the atom in its free atomic state. It is interesting to examine the impact of the gradient corrections for solid to solid reactions energies or solid to molecule reactions. In the Table II we have examined such a set of reactions for solid Al, In, As, and Sb reacting to form zincblende AlAs, AlSb, InAs, and InSb. It can be seen that the GC actually result in poorer agreement with experiment in all cases. We have examined several solid to molecule reactions and have found the same result, that is that the GC actually do more poorly than the LD in predicting reactions energies. We conclude therefore that the main contribution to the GC comes from the correction of the free atom energy, thus accounting for the improved agreement with experimental cohesive energies, but give poorer agreement in general for reactions among molecules and solids.

In Figs. 3-5 we have compared the GC and LD results for various elastic constants. For the bulk modulus we have done the calculation both at the calculated lattice constant and at the experiment lattice constant at 0K. There is some difficulty in obtaining reliable elastic constants for all materials, especially at zero temperature for which our calculations are done. Where they are available, we have compared to experimental values, extrapolated to zero temperature. For all of the elastic constants we see a nearly systematic improvement when the GC results are compared with the LD results. We find that in general the GC result in a softening of the bulk moduli relative to the LD, when the bulk moduli are evaluated at the calculated lattice constants. This can be attributed to the larger atomic volumes predicted by the GC. However if the bulk moduli are recalculated at the experimental atomic volumes, we find that the GC results in a systematic improvement over the LD.

¹ D. C. Langreth and M. J. Mehl, *Phys. Rev. Lett.* **47**, 446 (1981).

² U. van Barth and L. Hedin, *J. Phys. C* **5**, 1629 (1972).

³ M. Methfessel and M. van Schilfgaarde (unpublished).

⁴ M. Methfessel, *Phys. Rev. B* **38**, 1537 (1988).

⁵ R. Jones and A. Sayyesh, *J. Phys. C*, L653 (1986).

⁶ C. Filippi, D. J. Singh, and C. J. Umrigar, *Phys. Rev. B* **50**, 14947 (1994).

TABLES

TABLE I. Zero point energies for the tetrahedrally-bonded semiconductors. Calculation were done using a valence force field model and experimental elastic constants at zero temperature, where available, otherwise at room temperature

Material	Zero point energy (eV)
Si	0.05875
Ge	0.03311
AlP	0.05103
AlAs	0.04213
AlSb	0.03378
GaP	0.04321
GaAs	0.03161
GaSb	0.02551
InP	0.03398
InAs	0.02395
InSb	0.01981
ZnS	0.03606
ZnSe	0.02745
ZnTe	0.02291
CdTe	0.01731
HgS	0.02710
HgSe	0.01871
HgTe	0.01565

TABLE II. Compound binding energy relative to the elemental solid

Solid	Energy (eV)		
	LD	GC	exp
AlAs	0.516	0.412	0.605
AlSb	0.207	0.126	0.450
InAs	0.269	0.233	0.360
InSb	0.165	0.123	0.165

FIGURES

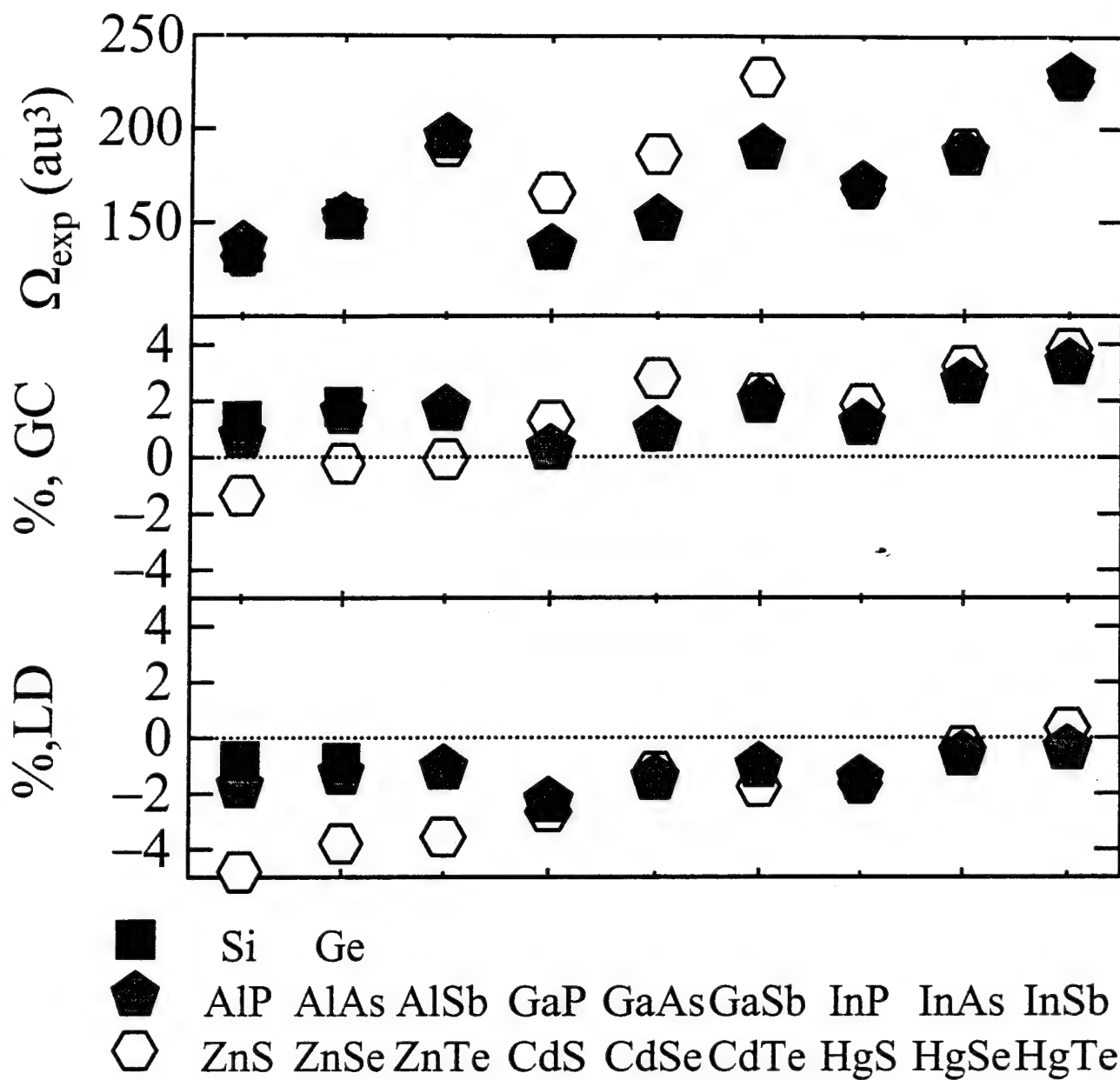
FIG. 1. In the upper panel the experimental volume per atom is plotted. In the middle panel the we have plotted $(\Omega_{GC} - \Omega_{exp})/\Omega_{exp}$, the percent error in the atomic volume when gradient corrections are used. In the lower panel we have plotted $(\Omega_{LD} - \Omega_{exp})/\Omega_{exp}$. (a) Elemental and compound semiconductors. (b) Elemental solids.

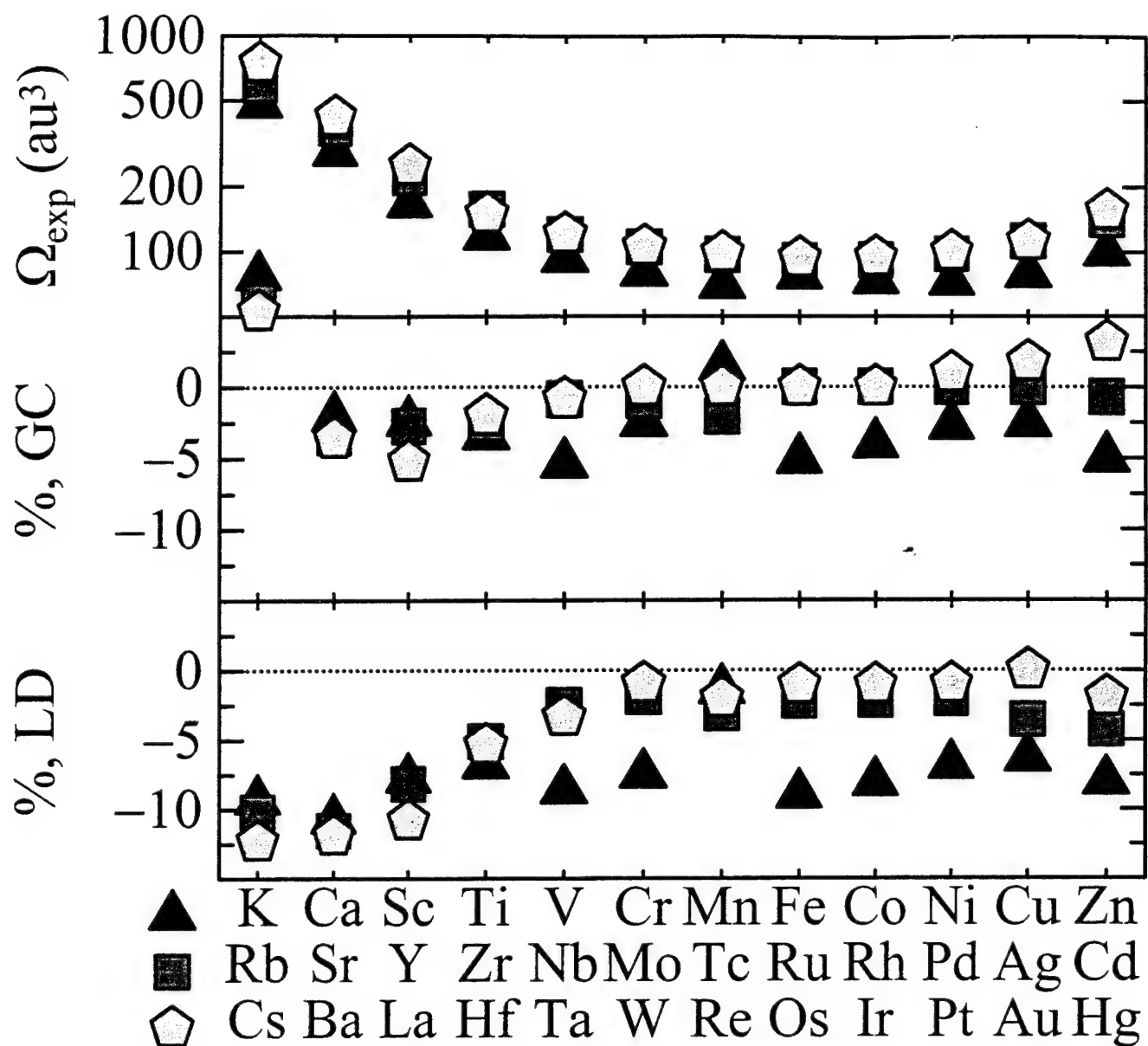
FIG. 2. In the upper panel the experimental cohesive energy per atom is plotted. In the middle panel the we have plotted difference $E_{coh}(GC) - E_{coh}(ex)$, in eV, and in the lower panel $E_{coh}(LD) - E_{coh}(ex)$, in eV. (a) Elemental and compound semiconductors. (b) Elemental solids.

FIG. 3. In the upper panel the experimental bulk moduli at zero temperature are plotted. In the middle panel the we have plotted difference percent error when gradient corrections are used $(B_{GC} - B_{exp})/B_{exp}$ and in the lower panel $(B_{LD} - B_{exp})/B_{exp}$, the percent error when local density is used. (a) Elemental and compound semiconductors. (b) Elemental solids.

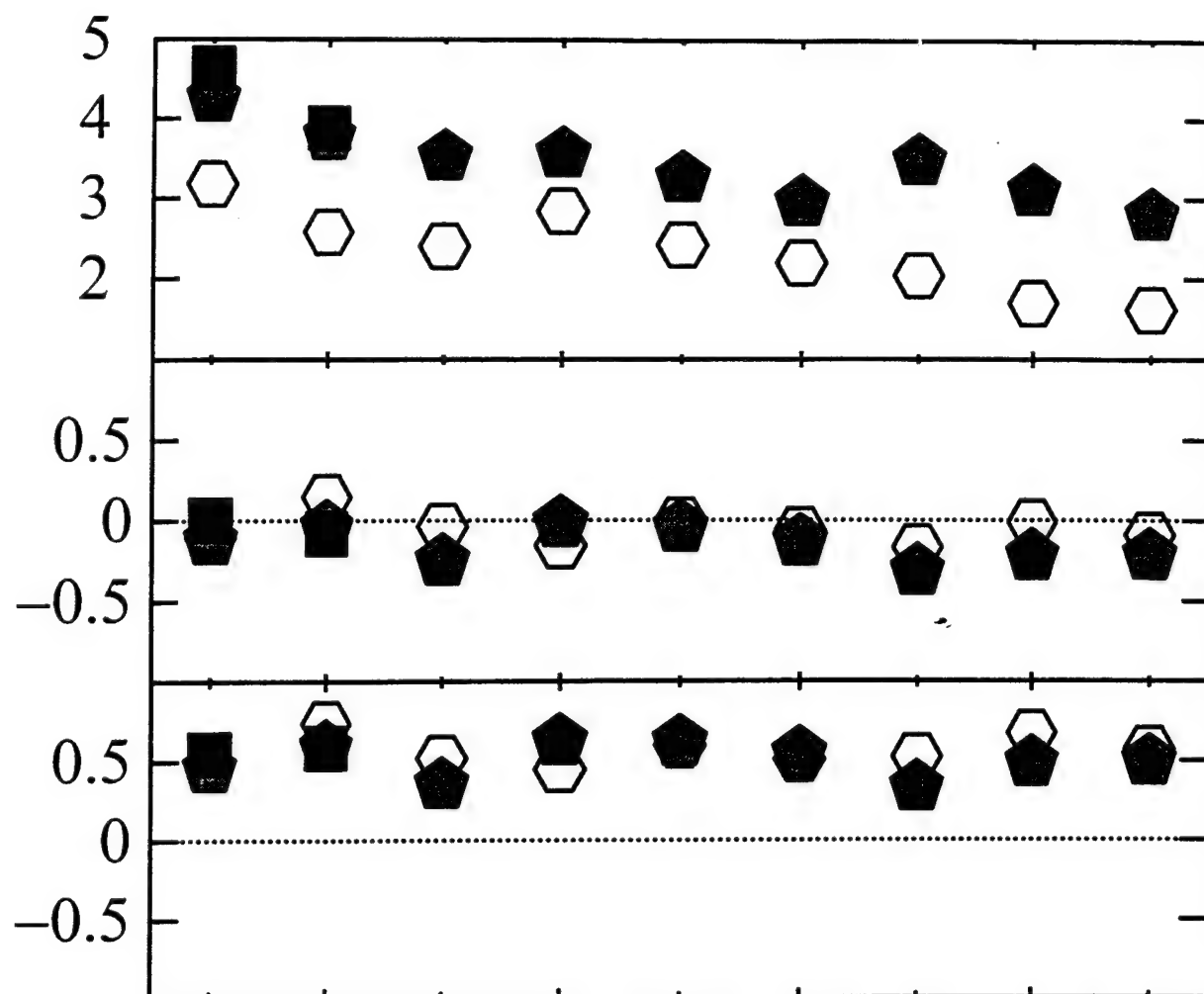
FIG. 4. In the upper panel the experimental values for distortions along the (a) $\langle 111 \rangle$ axis, C_{44} in the diamond and zincblende solids and (b) the c-axis, R are plotted. Units are in ergs/cm³. In the middle panel the we have plotted difference percent error when gradient corrections are used and in the lower panel the percent error when local density is used. (a) Elemental and compound semiconductors. (b) Elemental solids.

FIG. 5. In the upper panel the experimental values for the $C_{11} - C_{12}$ distortions, in ergs/cm³, are plotted for the diamond and zincblende solids. In the middle panel the we have plotted difference percent error when gradient corrections are used and in the lower panel the percent error when local density is used.

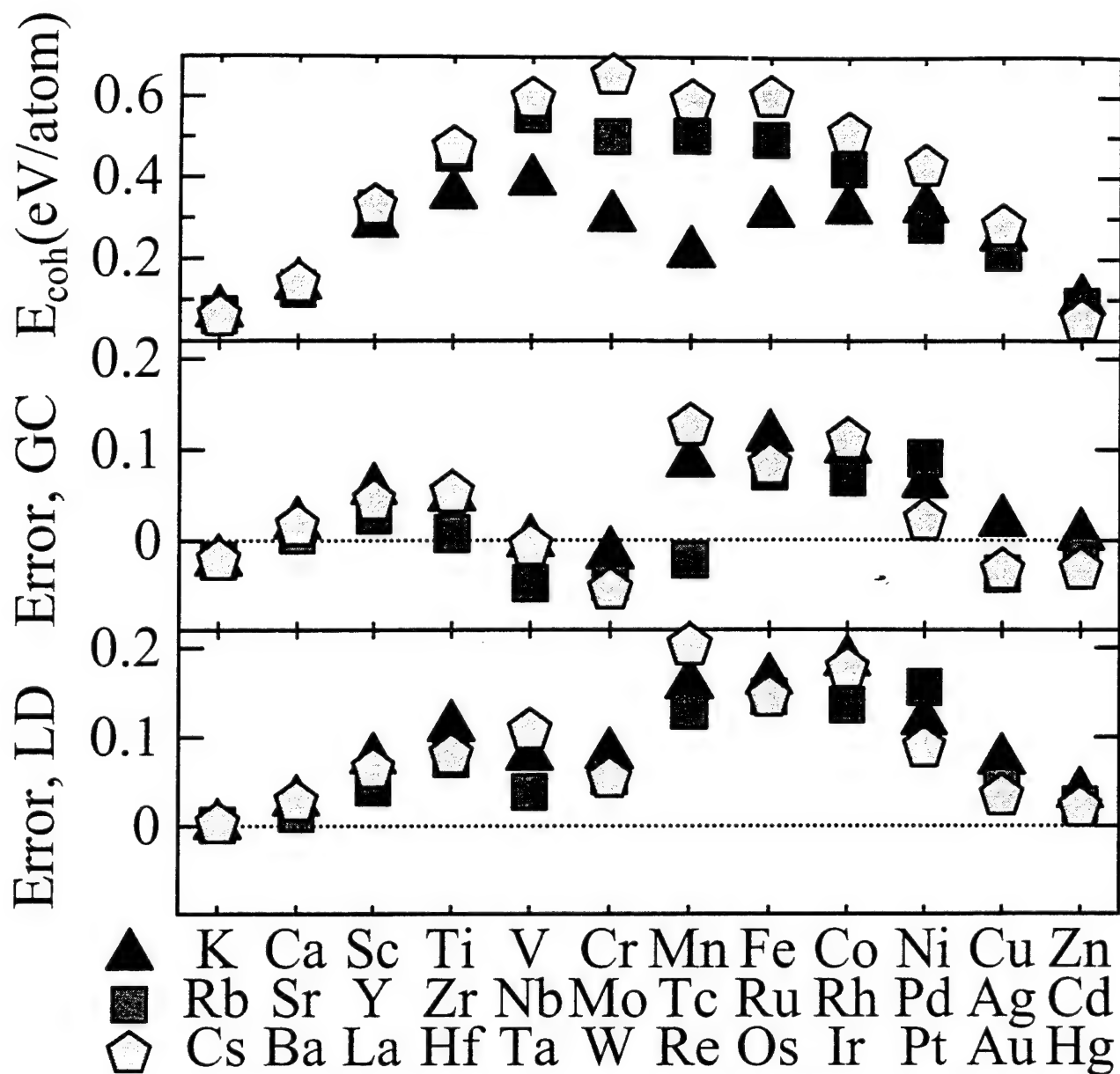


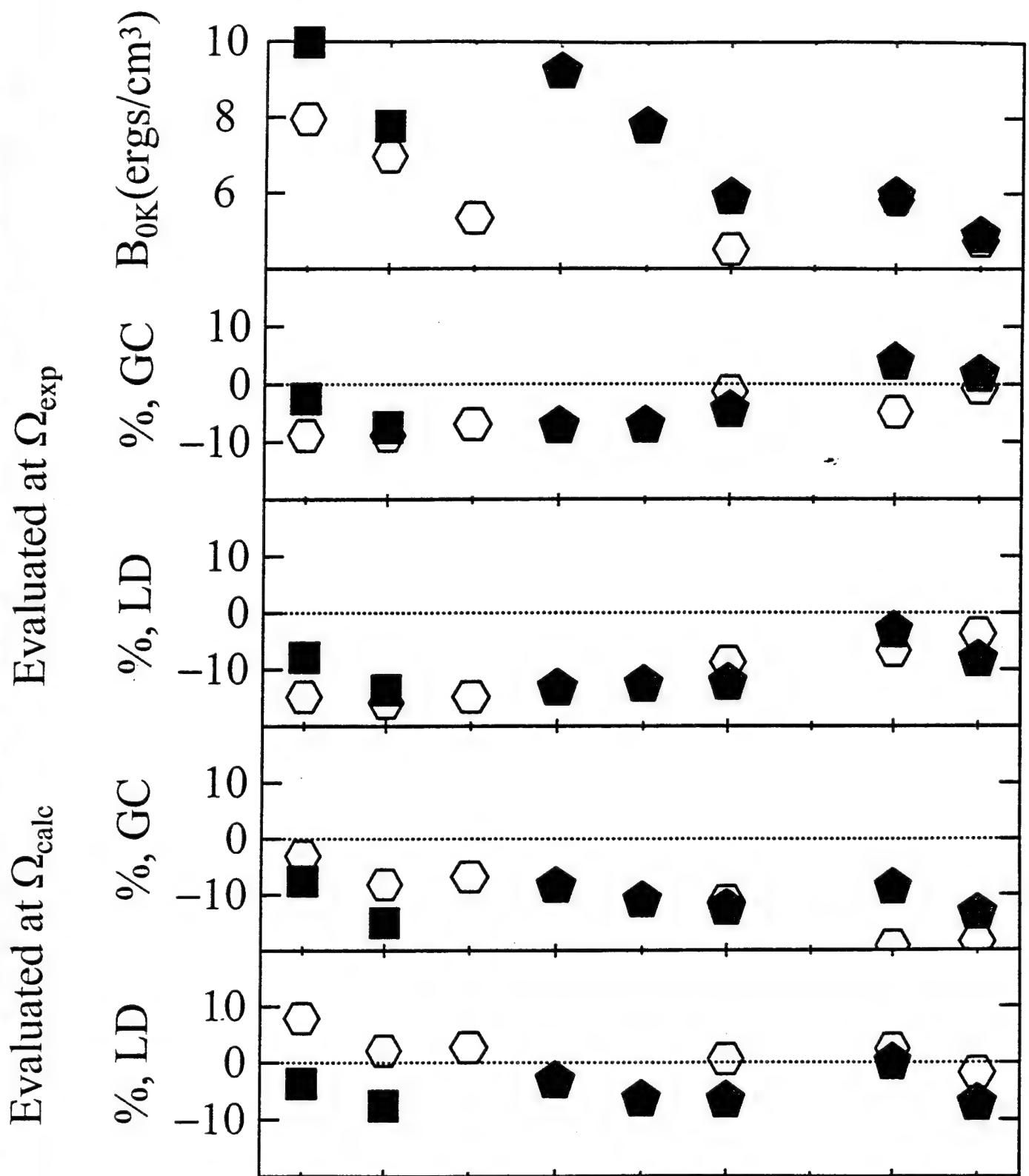


E_{coh} (eV/atom)
 Error, GC
 Error, LD

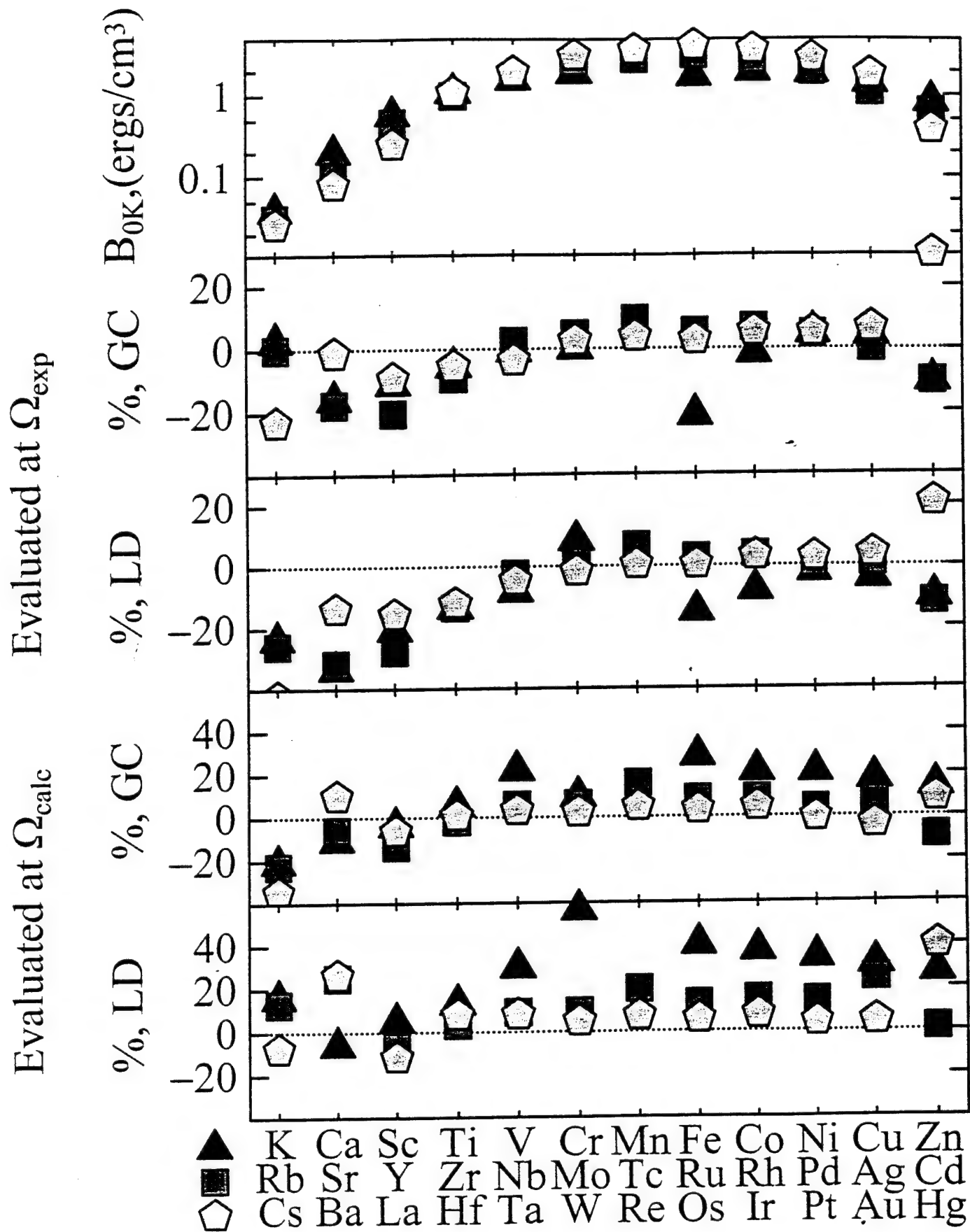


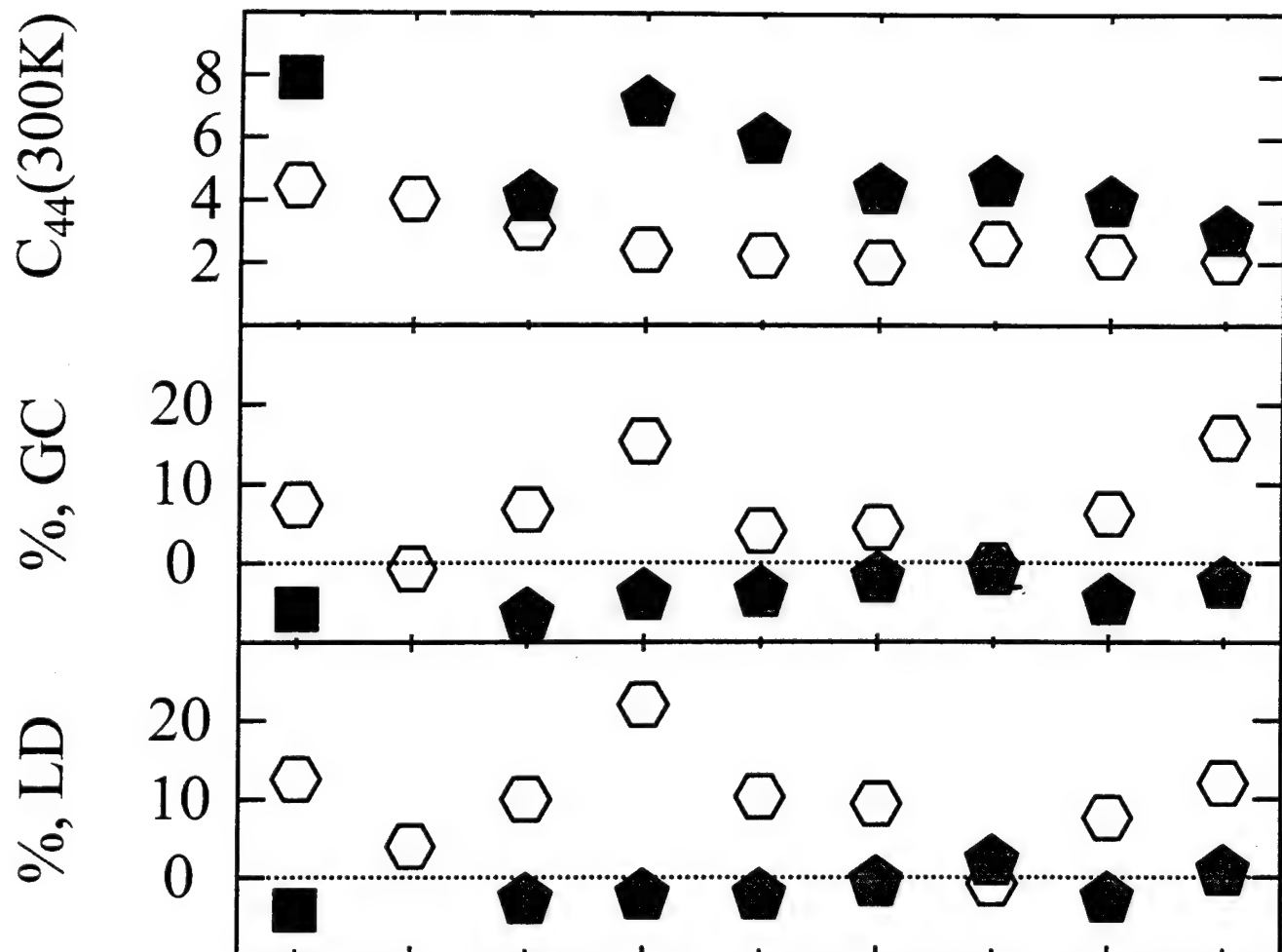
■ Si Ge
 ● AlP AlAs AlSb GaP GaAsGaSb InP InAs InSb
 ○ ZnS ZnSe ZnTe CdS CdSeCdTe HgS HgSeHgTe



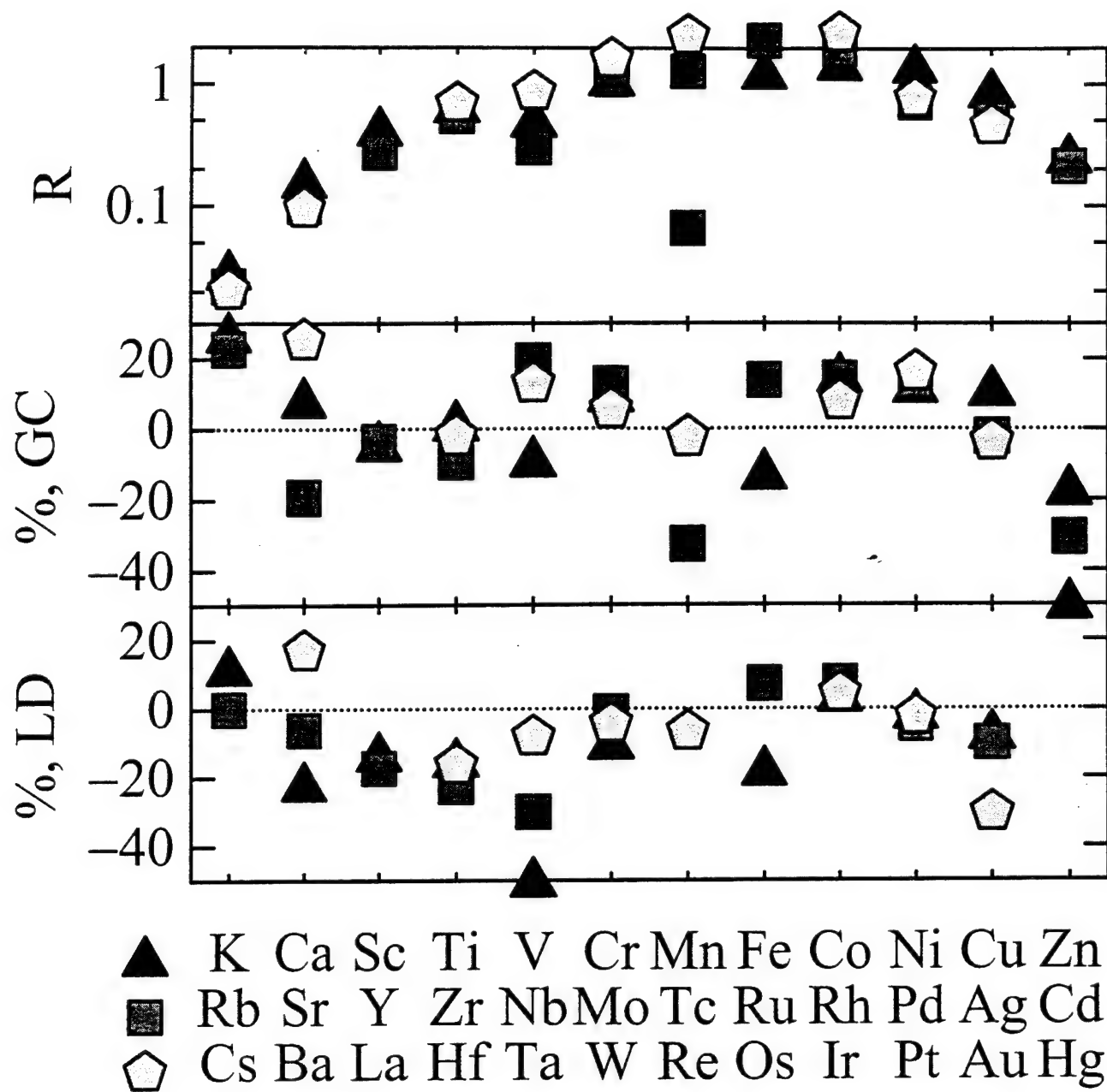


■ Si Ge
 ● AlP AlAs AlSb GaP GaAsGaSb InP InAs InSb
 ○ ZnS ZnSe ZnTe CdS CdSeCdTe HgS HgSeHgTe





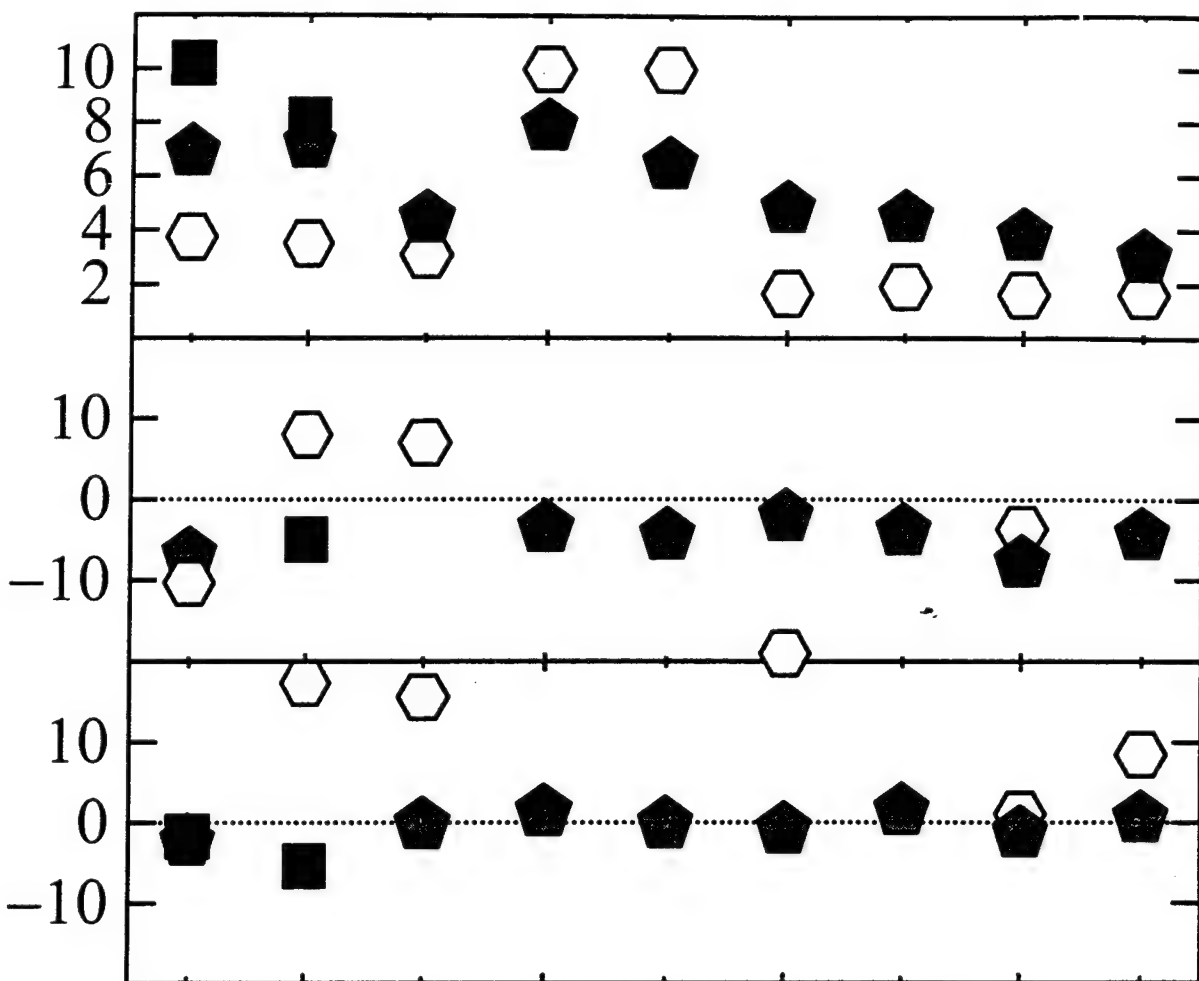
■ Si Ge
 ● AlP AlAs AlSb GaP GaAsGaSb InP InAs InSb
 ○ ZnS ZnSe ZnTe CdS CdSeCdTe HgS HgSeHgTe



$C_{11}-C_{12}(300K)$

%, GC

%, LD



Si Ge



AlP AlAs AlSb GaP GaAs GaSb InP InAs InSb



ZnS ZnSe ZnTe CdS CdSe CdTe HgS HgSe HgTe

APPENDIX M

Native point defects and degradation of ZnSe-based devices

M. A. Berding, M. van Schilfgaarde and App Sher

(In preparation for submission to *Appl. Phys. Lett.*)

DRAFT

Native point defects and degradation of ZnSe-based optical devices *

M. A. Berding, M. van Schilfgaarde, and A. Sher

SRI International, Menlo Park, California 94025

(September 6, 1995)

We propose a mechanism for the operational degradation of ZnSe-based device based on the generation of zinc Frenkel defects at stacking faults in the active layer. A first-principles method is used to compare the Frenkel defect formation energy in the bulk zincblende structure and in the wurtzite structure that is present at a stacking fault. In the wurtzite structure we find that the Frenkel defect formation energy is less than the recombination energy (E_g) of an electron-hole pair, while it is greater than this in the zincblende structure. Mid-gap levels of the zinc-vacancy associated with the Frenkel pair serve as additional non-radiative recombination sites; energy released from non-radiative recombination events also feed into the dislocation system, resulting in dislocation motion and multiplication. Eventually the supersaturation of zinc-interstitials the the stacking fault create metastable defects aligned along $\langle 100 \rangle$ directions, before being absorbed into the bulk zincblende lattice.

We have calculated the absolute density of native point defects in ZnSe using a self-consistent first-principles full-potential linearized muffin-tin orbital method and the local-density approximation to calculate the electronic contribution to the defect formation free energy. A gradient correction is added to the local-density result so that absolute reference to the chemical potential of the vapor phase can be made. Vibrational contributions to the defect free energy are calculated using a Green's function approach, based on a valence force field plus a point Coulomb model. Absolute equilibrium native defect concentrations as a function of constituent chemical potentials and the dopant concentration are predicted using a quasichemical formalism. We conclude that isolated native point defects are not responsible for the p-type doping limit found in ZnSe.

61:72.J, 61:72.C, 71:55.G, 42:70.K

00:00.X

I. INTRODUCTION

An important issue remaining in the development of a device technology of ZnSe-based light emitting diodes (LEDs) and lasers is the device degradation resulting in short operational lifetimes. The degraded devices exhibit large dark areas in the active region, where electron-hole recombination is dominated by nonradiative mechanisms¹⁻³. So called dark patches have been associated with the operational generation of networks of dislocations nucleated at threading dislocations and that penetrate the active region^{1,3}; dark spots have been associated with the original threading defects; and dark line defects (DLDs) along $\langle 100 \rangle$ directions that lie in the active region, that remain unidentified, but are not directly related to dislocations².

While portions of the degradation mechanism are understood, there has only been speculation concerning the underlying source of non-radiative sites and the origin of the DLDs. In this paper we propose an originating mechanism based on the excess generation of zinc-Frenkel defect pairs at stacking faults, where the formation energy is reduced to less than the recombination energy (E_g) of an electron-hole pair. Mid-gap levels of the zinc-interstitial associated with the Frenkel defect pair serve as additional non-radiative recombination sites. The energy released during non-radiative recombination events both generate additional Frenkel and feed into the dislocation system, resulting in dislocation motion and multiplication.

In Sec. II we outline the method we use to calculate the defect properties and densities in ZnSe. In Sec. III we summarize our results on the native point defect densities in ideal zincblende ZnSe. Our results on the Frenkel defect at the stacking fault are given in Sec. IV. Finally in Sec. V we have some concluding remarks.

* Submitted to *Physical Review B*, September 1995

II. METHODOLOGY

Total energy calculations for the defect reactions were all obtained with a full-potential version⁶ of the LMTO method, in the local density functional approximation of von Barth and Hedin.¹⁰ This method has been tested extensively for most of the elemental *sp* and *d* bonded solids, the II-VI, III-V and column IV semiconductors, and a host of other solids⁴.

In the FP-LMTO method, the only important approximation we make beyond the local-density approximation, lies in the treatment of the interstitial matrix elements. For the interstitial matrix elements we have used atom-centered basis, represented by Hankel functions consisting of a "triple kappa" basis 22 orbitals per atom, with energies -0.01 , -1 , and -2.3 Ry for the *s* and *p* orbitals, and -0.01 and -1 Ry for the *d*. The represent the electron density in the interstitial two Hankels per site and *lm* have been used. To ensure a good fit to the charge density and potential in the interstitial region of the zincblende solids, we include empty spheres at each tetrahedral interstitial site (rendering the sphere packing bcc for the ideal lattice). In addition, we added orbitals to the basis by centering them on the empty spheres. Both the charge density inside the spheres and the tails of Hankel functions centered on a neighboring sphere were expanded to $l = 6$. The core was allowed to relax during the self-consistency cycle. The semicore *d* electrons in the selenium were treated explicitly as valence states in a second panel; explicit treatment of these states was found to introduce a small but significant correction to the total energy. For the 16-atomic-site cells, the Brillouin zone integrals were done by a sampling method for the charge density, and the linear tetrahedron method for the band-structure energy, augmented by Blöchl weights, and a mesh of four divisions was used (13 k-points for the wurtzite structure and 8 k-points for the zincblende structure). For the 32-atom cells, a mesh of three divisions was used (7 k-points for the wurtzite structure and 4 for the zincblende structure). Further details of the method can be found in Berding *et al.*⁵ and references therein.

For consistency, unless otherwise noted, we have used the zinc atom in the vapor phase as our reference state for reporting energies. The defects we will consider, and their corresponding reactions are summarized in Table I. To calculate the total energy of a defect, supercells are used in which a periodic array of that defect is constructed. Defect formation energies are calculated by differences of cells with and without a defect with the ZnSe solid and the free zinc atom serving as the atom reservoir. Because we wish to calculate the formation energies in the dilute limit, we use the largest supercell computationally feasible. For this paper, all calculations were done using 16-atom supercells. Tests were done to check the convergence using 32-atom supercells; results are discussed below.

The local-density approximation generally overbinds the solids. Several systematic extensions of the local-density function have been proposed that are based on generalized gradient approximation for the exchange and correlation energies. We have considered one of these extensions, that proposed by Langreth and Mehl,⁷ and have examined the systematics in the gradient corrections to the lattice constants, cohesive energies, bulk modulus, and other elastic constants for a wide array of solids⁴. With few exceptions, the gradient correction systematically improves the prediction of the cohesive energy, although the prediction of the elastic constants often worsens slightly. The improvement in the cohesive energy is largely due to improvement in the calculation of the total energy of the free atoms, rather than the solid. Here, we are interested in calculating total energies for reactions in which a constituent is exchanged between the solid and the vapor, and thus the errors inherent in the local-density calculation of the cohesive energy will be present in these energies also. Because the relaxations do not change significantly when gradient corrections are added, we have completed the majority of the calculation, including the relaxation, within the FP-LMTO. The gradient correction energy, calculated at the LDA-determined relaxed positions, is then added to the LDA energy.

In general the lattice relaxes in the presence of a defect, thereby lowering the lattice energy. In the supercells we account for this relaxation by allowing the overall lattice constant of the supercell to relax to minimize the supercell total energy. Additionally, the nearest and second-nearest (for the 32-atom cell) neighbor atoms are allowed to relax radial about the defect site.

Relaxation energies are calculated only for the neutral defects and are assumed comparable in the ionized defects. Nonradial relaxations such as the trigonal and tetragonal distortions that split the degeneracy of the triply degenerate T_2 states may be important and may differ substantially for the different charge states of the system. Because the symmetry of the distortion depends on the charge state of the defect, distortions and charge states must be treated simultaneously. These distortions have not been considered in this paper.

The calculation of the ionization states of the defects is perhaps the most difficult part of the calculation of the native defect concentrations, in large part because of the inadequacies of the LDA in predicting the band gap of the semiconductors. Because the Coulomb fields associated with a defect may be extended, we expect that very large supercells will be needed to isolate the localized levels of an individual defect. The approach we use is similar to that discussed by Van de Walle *et al.*¹³ in which the shift in the Fermi level is examined as electrons are added to (or removed from) the defect, with a compensating uniform background charge added so as to maintain charge neutrality.

We have included in our calculation of the defect formation free energy a term that comes from modifications of the vibrational spectrum when a defect is introduced into the lattice. This change is especially important for reactions in which atoms are exchanged with the free zinc atom reservoir. We have used Keating's valence force-field model for the short-range elastic interactions.⁷ With a point-charge model to account for the Coulombic interactions,⁸ with a charge chosen to yield agreement with experiments for the zone center splitting of the transverse and longitudinal optical phonons. A Green's function approach is used to evaluate the lattice-defect induced modifications to the phonon spectrum; from the perturbed phonon density of states the change in the vibrational free energy can be calculated. Further details can be found in Berding *et al.*⁵.

Defect concentrations are calculated using the quasichemical analysis. Both neutral and ionized defects are included in the analysis, with the one electron donor and acceptor energies extracted from the LD calculations. Full Fermi-Dirac statistics are used, and appropriate degeneracies of the neutral and ionized states are included. Zinc free atoms are used as the reservoir and thus the zinc chemical potential is fixed by the zinc partial pressure. The free energy of this reservoir is included in the calculations. The chemical potential of the selenium is determined by the existence of the ZnSe in equilibrium with the zinc vapor phase via the relationship

$$\mu_{\text{Se}} + \mu_{\text{Zn}} = \mu_{\text{ZnSe}}. \quad (1)$$

Further details of the methodology can be found in Ref. 5.

III. NATIVE POINT DEFECT CONCENTRATIONS IN ZINCBLENDE BULK

Calculated defect formation energies for ZnSe are listed in Table I. One electron localized energy levels deduced from the LDA calculations are also listed in Table I. These energies were incorporated into the quasichemical analysis, and total defect concentrations as a function of zinc partial pressures over the existence region for undoped material were calculated (Fig. 1). We find the zinc vacancy is the dominant defect, except at the highest zinc pressures, for which the selenium vacancy dominates.

The calculations were redone, but with the an acceptor density of 10^{18} cm^{-3} , and with the acceptor energy level at 0.01 eV above the valence band edge. We find that isolated native point defects do not result in a substantial auto-compensation of the acceptor density. We have not addressed the absolute solubility of the acceptor in the host, as discussed by Van de Walle *et al.*¹⁴.

IV. FRENKEL DEFECT PAIR FORMATION ENERGY AT A STACKING FAULT

If we consider the $\langle 111 \rangle$ direction in the cubic zincblende crystal, the stacking sequence of the anion-cation bilayers is $\cdots ABCABC \cdots$ where A , B , and C indicate unique spatial positions of the bilayers in the xy -plane. The hexagonal wurtzite crystal can be interpreted as a simple alteration in the stacking to an $\cdots ABABAB \cdots$ sequence that maintains the four-fold tetrahedral coordination of each atom, but alters the crystal group properties. There are two types of simple stacking fault that can occur in the zincblende lattice. An extrinsic stacking fault can be interpreted as an extra plane added to the $\langle 111 \rangle$ stacking, altering the stacking to $\cdots ABCACBCABC \cdots$. An intrinsic stacking fault can be viewed as a bilayer removed from the $\langle 111 \rangle$ stacking, giving $\cdots ABCACABC \cdots$. Both of these classes of stacking faults can be viewed as a local region of wurtzite in an otherwise zincblende crystal.

To approximate the Frenkel defect formation energy at a stacking fault we consider the Frenkel defect in the wurtzite crystal structure. In the wurtzite crystal and at a stacking fault there are two inequivalent high-symmetry interstitial positions in which the nearest neighbor atoms are furthest away. We have only considered one class of interstitial, that which has the largest volume (defined as the volume of a sphere centered at the high symmetry interstitial site and with a radius defined by the distance to the first nearest neighbor). The interstitial volume at the stacking fault is 20% larger than the largest interstitial site in the zincblende lattice and thus we expect a lower interstitial formation energy at a stacking fault. The interstitial positions which we are modeling both at the stacking faults (wurtzite) and in the zincblende bulk are shown schematically in Fig. 2.

The formation energy for zinc Frenkel defect pair both in the zincblende and the wurtzite crystals are given in Table II. Note that for the neutral defect reactions, the zinc Frenkel defect formation energy is lowered 0.26 eV and 0.47 eV with respect to the zinc- and selenium- tetrahedral interstitial sites. Although the energy lowering is substantial, the Frenkel defect formation energy is still of the order of 6 eV and thus will be a relatively slow process for thermally generating zinc interstitials and zinc vacancies in ZnSe, and will not be a likely event as a consequence of a non-radiative recombination event.

Because there are localized levels associated with the zinc interstitial and vacancy, we must include the ionized levels in the analysis, as was done in Sec. III above. From Table I we see that the zinc vacancy is a double acceptor with both ionization levels at the valence band edge. The zinc interstitial (at a Se interstitial site) is a double donor with energy levels as 1.7 and 1.4 eV above the valence band edge. For Fermi energies below mid gap, that is in p -type regions, the electrons from the zinc interstitial donor levels will drop into the acceptor levels associated with the zinc vacancy. This self-compensation mechanism lowers the Frenkel defect pair formation by 3.1 eV for Zn_{ISe} and to ~ 2.8 eV for the zinc interstitial in the wurtzite structure. Although the Frenkel defect formation energy in the zincblende is still larger than the band gap, in the wurtzite crystal, and therefore at a stacking fault, the energy is less than that released by a non-radiative e - h recombination event. Because the binding energy of the zinc interstitial is higher at a stacking fault, they will be trapped there. The mid-gap levels associated with the zinc interstitial will most likely serve as additional non-radiative recombination centers.

Based on our findings, we propose a degradation mechanism in which excessive Frenkel defect pairs are generated at stacking faults present in QW structure via the energy available through non-radiative recombination events. The zinc interstitials generated in these events serve as additional nonradiative recombination centers, thus amplifying the Frenkel defect generation rate. The thermal energy made available from non-radiative recombination events at the Frenkel defect-dressed stacking faults produce temperature gradients that could also contribute to dislocation motion and multiplication.

This model may also provide the basis of an explanation for the so called dark line defects. Some of the zinc interstitials which are formed at the stacking fault should be trapped there in potential wells, but others may diffuse away. In the presence of all of the dislocation motion and multiplication, one may have an ordering of the interstitials along the $\langle 100 \rangle$ directions. Although we have not completed this portion of our modeling, certainly our model of Frenkel defect generation at the stacking fault provides a explanation of a source of very mobile point defects which may be correlated with these dark line defects.

V. SUMMARY AND CONCLUSIONS

We have shown that isolated native point defects are not responsible for the observed p -type doping limits in ZnSe. Based on the lower zinc Frenkel defect formation energy at a stacking fault, we have propose a mechanism for the operational optical degradation of ZnSe devices. In the wurtzite structure we find that the Frenkel defect formation energy is less than the recombination energy (E_g) of an electron-hole pair, while it is greater than this in the zincblende structure. Mid gap levels of the zinc-vacancy associated with the Frenkel pair serve as additional non-radiative recombination sites; energy relased from non-radiative recombination events also feed into the dislocation system, resulting in dislocation motion and multiplication.

FIG. 1. Native point defect concentrations over the existence region of ZnSe at 500° C

FIG. 2. Intrinsic stacking fault and location of interstitial and bulk zincblende interstitials.

TABLE I. Defect reactions considered for compound ZnSe. Notation is as follows: The primary symbol refers to the species, the subscript refers to the site that the species occupies, with no subscript indicating that the species is occupying its usual lattice site. V corresponds to a vacancy, I an interstitial, $free$ is the free atom reference state. Two types of interstitials are considered, both occupying tetrahedral sites, the first surrounded by four zinc nearest neighbors, I_{Zn} , the second surrounded by four selenium nearest neighbors, I_{Se} . All species are neutral. Formation energies are calculated using 16-atom supercells. Localized levels are all with respect to the valence band edge, and are calculated using 54-atom supercells and the atomic spheres approximation.

Defect	Reaction			Energy (eV)			E_d (eV)	E_a (eV)
				LD	GC	Total		
V_{Zn}	ZnSe	\rightarrow	$V_{Zn}Se + Zn_{free}$	5.09	-0.46	4.63		
V_{Se}	Zn_{free}	\rightarrow	ZnV_{Se}	0.72	0.32	1.04		
$ZnZn$	$2Zn_{free}$	\rightarrow	$ZnZn_{Se}$	-1.37	1.00	-0.37		
$SeZn$	$2ZnSe$	\rightarrow	$SeZn_{Se} + 2Zn_{free}$	9.57	-1.16	8.42		
ZnI_{Zn}	Zn_{free}	\rightarrow	ZnI_{Zn}	1.10	0.32	1.42		
ZnI_{Se}	Zn_{free}	\rightarrow	ZnI_{Se}	1.36	0.26	1.62		
SeI_{Zn}	ZnSe	\rightarrow	$SeI_{Zn} + Zn_{free}$	7.55	-0.75	6.80		
SeI_{Se}	ZnSe	\rightarrow	$SeI_{Se} + Zn_{free}$	9.10	-0.85	8.24		

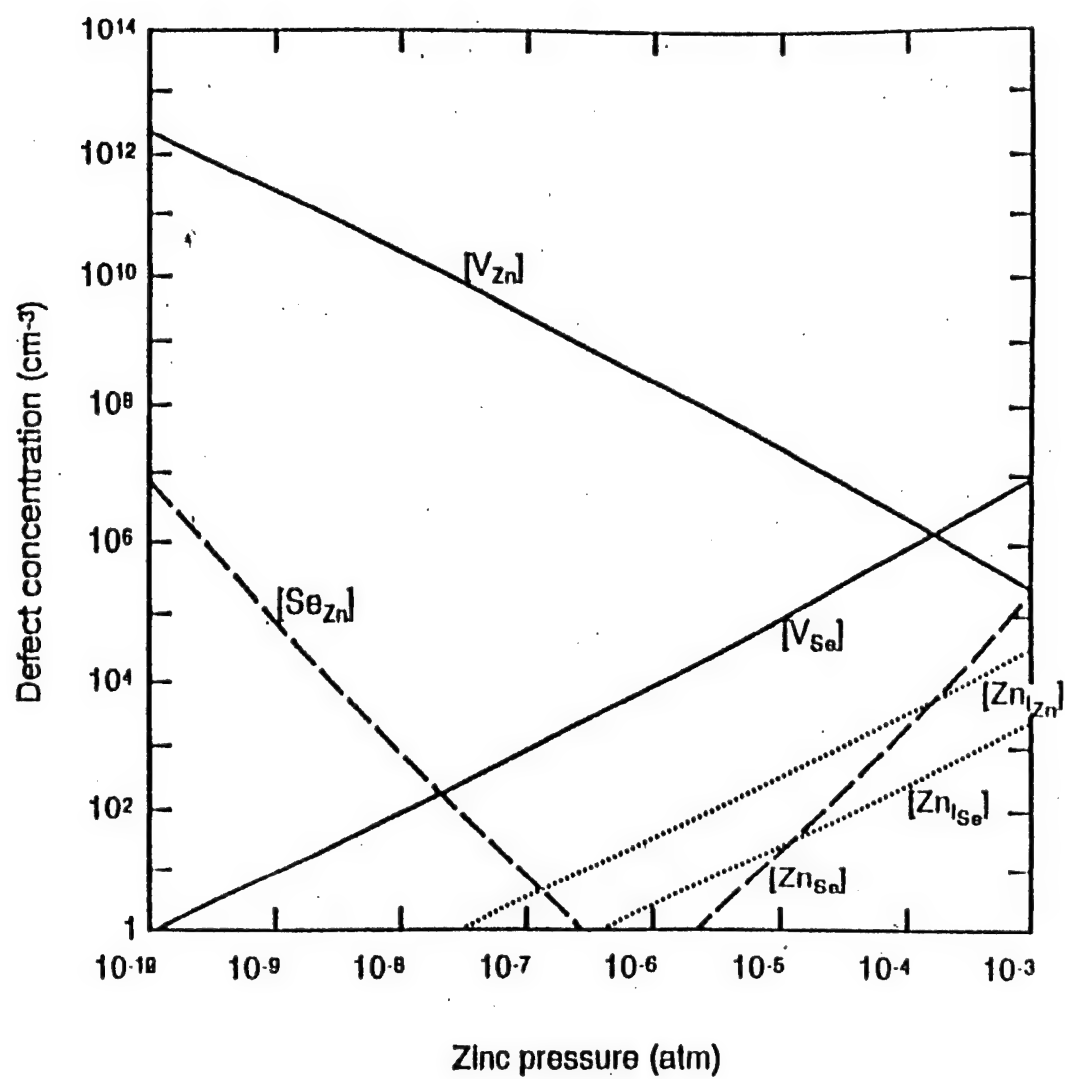
TABLE II. Frenkel defect reactions

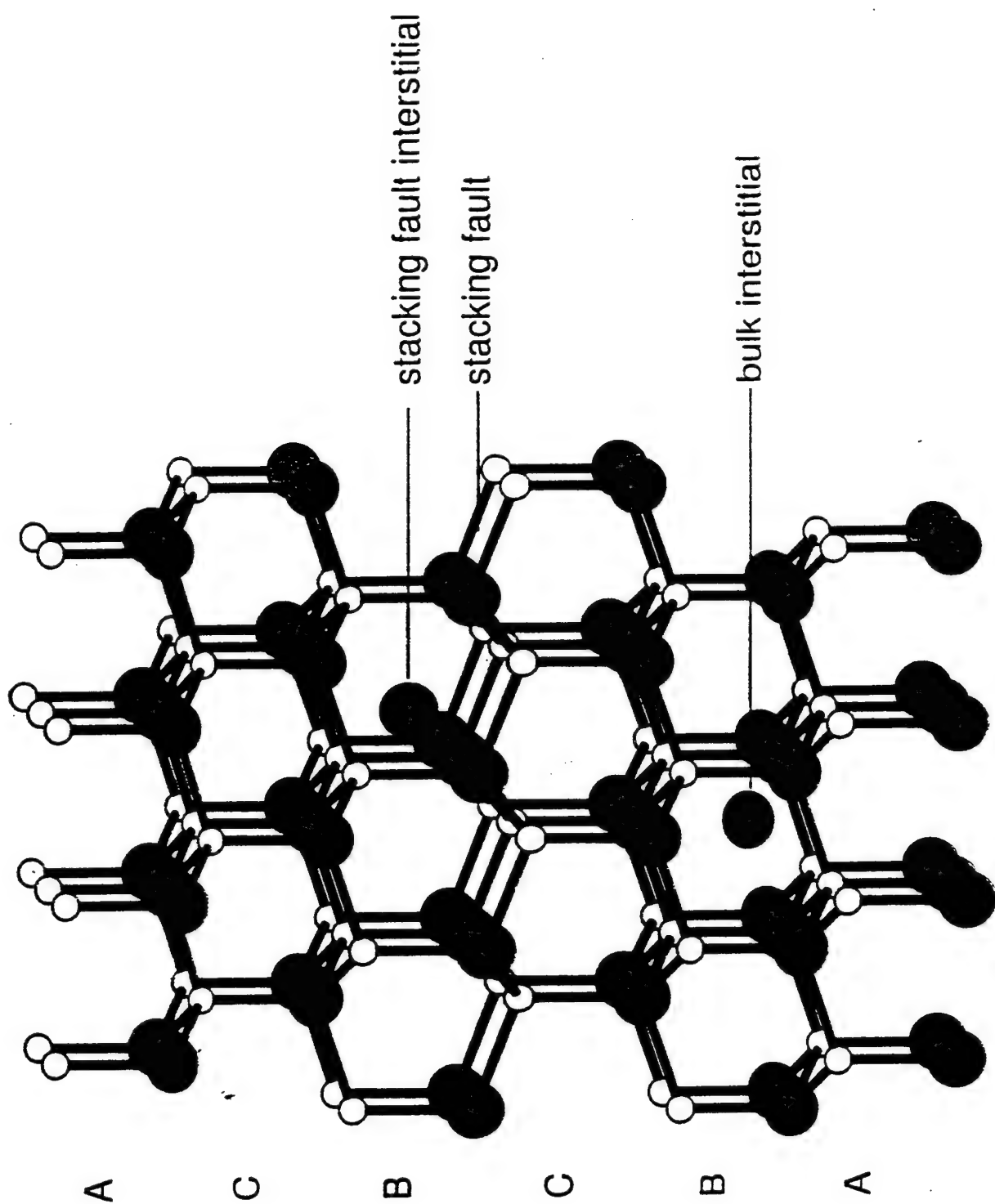
Host	Reaction			Energy (eV)
ZB	$ZnZn$	\rightarrow	$V + ZnI_{Zn}$	6.05
ZB	$ZnZn$	\rightarrow	$V + ZnI_{Se}$	6.25
WZ	$ZnZn$	\rightarrow	$V + ZnI^{\bullet}$	5.78
ZB	$ZnZn$	\rightarrow	$V^{--} + ZnI_{Zn}^{++}$	3.1
WZ	$ZnZn$	\rightarrow	$V^{--} + ZnI_{Zn}^{++}$	2.8

ACKNOWLEDGMENTS

We wish to acknowledge the work of Michael Methfessel in development of the first-principles codes applied in this work. This work has been supported ARPA Contract MDA972-92-C-0053.

- ¹ S. Guha, J. M. DePuydt, M. A. Haase, J. Qiu, and H. Cheng, *Appl. Phys. Lett.* **63**, 3107 (1993).
- ² S. Guha, H. Cheng, M. A. Haase, J. M. DePuydt, J. Qiu, B. J. Wu, and G. E. Hofer, *Appl. Phys. Lett.* **65**, 801 (1994).
- ³ G.C. Hua, N. Otsuka, D. C. Grillo, Y. Fan, J. Han, M. D. Ringle, R. L. Gunshor, M. Hovinen, and A. V. Nurmikko, *Appl. Phys. Lett.* **65**, 1331 (1994).
- ⁴ M. van Schilfgaarde, M. A. Berding, and A. T. Paxton, (submitted to *Phys. Rev. B*).
- ⁵ M. A. Berding, M. van Schilfgaarde, and A. Sher, *Phys. Rev. B* **50**, 1519 (1994).
- ⁶ M. Methfessel and M van Schilfgaarde (unpublished).
- ⁷ P. N. Keating, *Phys. Rev.* **145**, 637 (1966).
- ⁸ A. A. Maradudin, E. W. Montroll, G. H. Weiss, and I. P. Ipatova, in *Solid State Physics Supp. 3*, edited by H. Ehrenreich, F. Seitz and D. Turnbull (Academic Press, New York 1971), p. 1.
- ⁹ M. Methfessel, *Phys. Rev. B* **38**, 1537 (1988).
- ¹⁰ U. von Barth and L. Hedin, *J. Phys.* **C5**, 1629 (1972).
- ¹¹ D. B. Laks, C. G. Van de Walle, G.F. Neumark, and S. T. Pantelides, *Phys. Rev. Lett.* **66**, 648 (1991).
- ¹² D. B. Laks, C. G. Van de Walle, G.F. Neumark, P. E. Blöchl, and S. T. Pantelides, *Phys. Rev. B* **45**, 10965 (1992).
- ¹³ C. G. Van de Walle, P.J. H. Denteneer, Y. Bar-Yam, and S. T. Pantelides, *Phys. Rev. B* **39**, 10791 (1989).
- ¹⁴ C. G. Van de Walle, D. B. Laks, G.F. Neumark, and S. T. Pantelides, *Phys. Rev. B* **47**, 9425 (1993).
- ¹⁵ D. J. Chadi, *Phys. Rev. Lett.* **72**, 534 (1994).





APPENDIX N
REPORT FROM SUBCONTRACT TO PROF. A.-B. CHEN
AT AUBURN UNIVERSITY

Professor A.-B. Chen contributed to many of the results and publications referred to in the main body of this report. Below are several additional contributions made under the subcontract to Auburn University.

N-1. Systematic calculation of band structures of semiconductors and alloys.

A hybrid empirical tight-binding and pseudo-potential (HPT) method has been used for a systematic calculation of band structures of semiconductors and their alloys. This method not only produces accurate band structures such as the valence bands, the band gaps, the effective masses and the spin-orbit splitting with an accuracy comparable to that of the non-local empirical pseudo-potential method, but also enables the execution of alloy calculation using the multiple-scattering Green function. This method has been successfully applied to a collection of III-V and II-VI compounds and their alloys, especially including ZnTe, CdTe, HgTe, ZnSe, CdSe, HgSe and their narrow-gap alloys $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$, $\text{Hg}_{1-x}\text{Zn}_x\text{Te}$, $\text{HgSe}_{1-x}\text{Te}_x$, $\text{Hg}_{1-x}\text{Cd}_x\text{Se}$, and $\text{Hg}_{1-x}\text{Zn}_x\text{Se}$. These results were reported at the 1994 APS March Meeting. Detailed results are available from the book entitled "Semiconductor Alloys: Physics and Materials Engineering", by A.-B. Chen and A. Sher, to be published by Plenum Press (1995).

N-2. Molecular Dynamics Study of Disordered Semiconductor Alloys

To study disordered systems and non-equilibrium processes, we have started our molecular dynamics (MD) simulation research. We have carried out MD simulations for random semiconductor alloys containing 64 to 1000 atoms per unit cell using a valence-force-field (VFF) model to study the dependence of bond lengths and excess energies on the number of atoms (N) simulated. We found that the configurational fluctuations in both quantities decrease rapidly with an increase in N. For example, the first-neighbor bond length fluctuation in $\text{Ga}_{1-x}\text{In}_x\text{As}$ at $x=0.5$ decreases from 0.014\AA to 0.004\AA as N increases from 64 to 512. At the time the excess energy fluctuation is reduced from 10 meV to 2 meV. However, the values averaged over a number of configurations converge very fast with N. We conclude from this study that the structural properties of a random zincblende alloy can be obtained from a small N ($N=64$) and a dozen or so configurations. This result is then used in systematic MD simulation of semiconductor alloys. These results are being written for publication.

N-3. Monte-Carlo Simulation of Phase Diagrams of Semiconductor Alloys

Directly using the HPT Hamiltonian in Monte-Carlo (MC) simulations, one can avoid the uncertainties introduced in conventional phase diagram calculations based

on cluster energies and approximate entropy expressions. However, the extrapolation from finite sizes to infinite bulk is still not well established. We have tested a simple formula that extrapolates the average energy from finite size to infinite for an alloy at a given temperature T and concentration x . These extrapolated $E(x,T)$ are then used to calculate the Helmholtz free energy $F(x,T)$ and the phase diagram. Our test on the exactly solvable Ising model has yielded a reasonably accurate solid solution phase diagram (miscibility gaps as function of temperature). This procedure is being used in the MC simulation of phase diagrams of ternary semiconductor alloys using a tight-binding model.

N-4. Density Matrix and Order-N Algorithm

Although the current LDA techniques allow calculations to be formed for a modest size with N equal about 100, most practical materials problems required larger calculations with N at least one or two orders larger than the current limit. The limiting factor is in the diagonalization of the single-particle Hamiltonian, which usually takes CPU time proportional to cubic power of N . A method based on the density matrix has been proposed that reduces the calculational steps from the third power of N to linear. We have devised a simple and effective iterative method to obtain the density matrix. Our test on a tight-binding model has confirmed the validity of this method. For example, Table N-1 shows the total electronic energies per unit cell for GaAs. With a truncation in the density matrix to second neighbors and with 6 iterations, the method produces an energy accurate to 1%. We are extending this test to other structural properties such as the lattice parameter and elastic constants.

Table N-1 Total electronic energies (in eV) per unit cell calculated from the density matrix in the first few iterations. The first and the second rows correspond to density matrices that are truncated at the first and second neighbors respectively.

TRUNCATION	Iteration						
	1	2	3	4	5	6	exact
1st	-21.637	-26.531	-29.916	-29.713	-30.011	-30.021	-31.396
2nd	-21.637	-26.530	-30.317	-30.721	-31.042	-31.093	

N-5. Shallow Donor Levels with Anisotropic Masses

A general method for finding the energies and wave functions for anisotropic masses and potentials has been studied. The method uses wavefunctions expanded in spherical harmonics. The coupled differential equations for the radial components of each basis are integrated numerically. The energy eigenfunctions and their radial derivatives are then matched at a spherical boundary. Using this method, we have obtained energy levels for shallow donors with threefold mass anisotropy. The detailed results have been published in *Phys. Rev. B* **48**, 8541 (1993), R.S. Pfeiffer, Y.J. Huang, and A.-B. Chen. This paper is included as Appendix O.

APPENDIX O

Bound states for anisotropic potentials and masses

R. S. Pfeiffer, Y.-J. Huang and A.-B. Chen

Phys. Rev. B **48**, 8541 (1993)

PHYSICAL REVIEW B

CONDENSED MATTER

THIRD SERIES, VOLUME 48, NUMBER 12

15 SEPTEMBER 1993-II

Bound states for anisotropic potentials and masses

R. S. Pfeiffer, Y.-J. Huang, and A.-B. Chen

Department of Physics, Auburn University, Auburn, Alabama 36849-5311

(Received 13 April 1993)

A general method for finding the energies and wave functions for anisotropic masses and potentials has been studied. This method uses basis wave functions expanded in spherical harmonics. The coupled differential equations for the radial components of each basis function are integrated numerically. The energy eigenfunctions and their radial derivatives are then matched at a spherical boundary. Special treatment is needed to ensure linear independence of the basis functions at the boundary. The applicability of this method and the speed of convergence are tested on anisotropic harmonic oscillators. The method is then applied to the Coulomb potential with anisotropic masses. With a basis of five or fewer spherical harmonics, our method produces energies which converge to values lower than those previously reported. We have also obtained energy levels for the Coulomb potential with threefold mass anisotropy. This method should be applicable to other anisotropic problems with a single potential minimum. In particular, it should facilitate the employment of full-potential Green-function band theory.

I. INTRODUCTION

We here describe a method for calculating the energies and wave functions of bound states in anisotropic potentials and present results for the Coulomb potential with three anisotropic masses. Our interest in this problem was suggested by two more general problems: donor levels in complex semiconductors and Green-function band theory for nonspherical potentials.¹ As the structure of semiconductors becomes more complicated, the symmetry of the state at the conduction-band minima may be reduced and the effective-mass tensor becomes more complex. For example, in SiC, on going from the 3C (cubic or zinc blende) and 2H (hexagonal or wurtzite) structures to higher α polytypes, such as 4H and 6H, the effective-mass anisotropy goes from twofold to threefold.² Calculation of donor levels with three unequal masses, even in the hydrogenlike potential model, is not a trivial problem, and, as far as we know, has not been reported in the literature. The problem can be converted into one with an isotropic mass but anisotropic potential. We note in this connection that treating nonspherical potentials has been a central issue in extending the Korringa-Kohn-Rostoker^{3,4} (KKR) Green-function band theory⁵ beyond the muffin-tin approximation. By working on a single potential, we can isolate one of the important problems in the full-potential KKR theory—the construction of the basis functions.

So far, calculations with mass anisotropy have dealt

only with twofold anisotropy, which can be characterized by a single parameter α , ranging from 0 to 1. $\alpha=0$ is the isotropic case, $\alpha=1$ the two-dimensional (circular symmetry) case. The best calculations to date for mass anisotropy in a Coulomb potential for a wide range of values of α are those of Faulkner⁶ for the ground and several excited states and of Pollman⁷ for the ground state. Faulkner employed a basis of 18 hydrogen wave functions, consisting of 6 radial functions for each of the three chosen values of the angular momentum, while Pollman employed a parametrized product of the exact solutions for $\alpha=0$ and 1. Both sets of calculations deal entirely with twofold anisotropy. More recent reports^{8,9} tend to focus on particular semiconductors, and therefore single values of α . Our results for three unequal masses should assist future experimental analysis of donor levels in semiconductors with this type of mass anisotropy.

The rest of this paper is arranged as follows. Section II describes the method. We present the results in Sec. III, which includes tests of the convergence in an anisotropic harmonic oscillator, a comparison of our results with previous calculations for the Coulomb potential with twofold mass anisotropy, and our results with threefold mass anisotropy. The final section, Sec. IV, is a summary and discussion.

II. METHOD

We begin by assuming an anisotropic potential $V(\mathbf{r})$. For the case of an isotropic potential and anisotropic

mass, the kinetic energy in \mathbf{k} space is taken to be $T = (\hbar^2/2)(k_x^2/m_1 + k_y^2/m_2 + k_z^2/m_3)$. Transformation to new coordinates $x' = x\sqrt{m_1/m}$, $y' = y\sqrt{m_2/m}$, and $z' = z\sqrt{m_3/m}$, where $1/m = \frac{1}{2}(1/m_1 + 1/m_2)$, produces an isotropic kinetic energy with an anisotropic potential.

The way the bound states are calculated is similar to that widely used for a spherical potential: Draw a sphere of radius r_0 , construct a wave function inside the sphere ψ_{in} , and an outside wave function ψ_{out} by numerical integration, and match the value and derivative of the two wave functions at $r = r_0$. Of course ψ_{in} and ψ_{out} have to be regular functions for $r \rightarrow 0$ and $r \rightarrow \infty$, respectively. In the actual calculation, the eigenfunctions are expanded in basis functions:

$$\psi_{in} = \sum_{\gamma} a_{\gamma} U^{\gamma}, \quad (1a)$$

$$\psi_{out} = \sum_{\beta} b_{\beta} W^{\beta}. \quad (1b)$$

The basis functions U^{γ} and W^{β} are further expanded in real spherical harmonics Y_{lm} . We describe the procedure for constructing the U^{γ} below. A similar procedure is used to construct the W^{β} .

We truncate the expansion at a maximum value of l , denoted l_{max} . For a given symmetry, U^{α} is expanded as

$$U^{\gamma}(\mathbf{r}) = \sum u_{lm}^{\gamma}(r) Y_{lm}(\theta, \phi), \quad (2)$$

where the summation includes all allowed values of l and m for $l \leq l_{max}$. If the total number of allowed values of l and m is N , then N independent basis functions U^{γ} need to be constructed.

For a specific γ , the radial functions u_{lm}^{γ} are coupled through N (one for each lm) differential equations:

$$\left[-\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d}{dr} \right) + \frac{l(l+1)}{r^2} - E \right] u_{lm}^{\gamma}(r) + \sum_{l'm'} V_{lm'l'm'}(r) u_{l'm'}^{\gamma}(r) = 0, \quad (3)$$

where E is an energy parameter and the potential matrix elements are defined as

$$V_{lm'l'm'}(r) = \int Y_{lm}(\theta, \phi) V(r, \theta, \phi) Y_{l'm'}(\theta, \phi) d\Omega. \quad (4)$$

Equation (3) can be treated as a matrix differential equation for the vector $u^{\gamma}(r)$, which is integrated outward by the matrix Numerov method, with the values of the vector being specified at the two initial points. We obtain the other basis functions similarly by specifying other initial vectors, the entire group of vectors being mutually independent. We have found it necessary, however, to reorthogonalize the set of N vectors u^{γ} at several different values of r as we integrate outward whenever $N > 2$. If this is not done, the vectors lose their independence by the time r_0 is reached.

After the radial functions u_{lm}^{γ} and w_{lm}^{β} are obtained, we seek to match the total wave functions (i.e., $\psi_{in} = \psi_{out}$) at $r = r_0$. For the wave functions to match at all angles, the radial term associated with each spherical harmonic must match independently. Thus, for every lm , we have

$$\sum_{\gamma} a_{\gamma} u_{lm}^{\gamma}(r_0) = \sum_{\beta} b_{\beta} w_{lm}^{\beta}(r_0). \quad (5)$$

Similarly, continuity of the first derivative $\dot{\psi}_{in} = \dot{\psi}_{out}$ at $r = r_0$ produces

$$\sum_{\gamma} a_{\gamma} \dot{u}_{lm}^{\gamma}(r_0) = \sum_{\beta} b_{\beta} \dot{w}_{lm}^{\beta}(r_0) \quad (6)$$

for every lm .

Equations (5) and (6) together constitute a set of $2N$ linear equations which can only have solutions at allowed energies of the total wave function. The determinant or nearest-to-zero eigenvalue of the matrix constructed from the various $u_{lm}^{\gamma}(r_0)$ and $w_{lm}^{\beta}(r_0)$ and their first derivatives is evaluated as a function of the energy E and the roots found by a numerical search.

III. APPLICATION AND RESULTS

The method just described is applied to the anisotropic harmonic oscillator and the Coulomb potential with anisotropic masses in this section.

In the hydrogenic approximation to the donor problem, the wave equation for the donor electron takes the form

$$-\left[\frac{1}{m_1} \frac{\partial^2}{\partial x^2} + \frac{1}{m_2} \frac{\partial^2}{\partial y^2} + \frac{1}{m_3} \frac{\partial^2}{\partial z^2} + \frac{2}{\kappa r} \right] \psi = E \psi, \quad (7)$$

where the electron mass and the Rydberg are taken to be unity, κ is the dielectric constant of the bulk semiconductor, and m_1 , m_2 , and m_3 are the effective mass components expressed as multiples of the electron mass. The coordinate transformation described at the beginning of Sec. II, together with an effective mass m defined by $1/m = \frac{1}{2}(1/m_1 + 1/m_2)$ and an energy scaled by $E = (m/\kappa^2)\epsilon$, produces the equation employed in the actual calculations:

$$-\nabla^2 \psi - \frac{2}{\sqrt{(1+\alpha_2)x^2 + (1-\alpha_2)y^2 + (1-\alpha_1)z^2}} \psi = \epsilon \psi. \quad (8)$$

In terms of the original masses, the anisotropy parameters α_1 and α_2 are

$$\alpha_1 = 1 - \frac{m}{m_3}, \quad (9a)$$

$$\alpha_2 = \frac{m_2 - m_1}{m_2 + m_1}. \quad (9b)$$

The isotropic case corresponds to $\alpha_1 = \alpha_2 = 0$. The solutions to Eq. (8) are then the hydrogen wave functions and have energies $\epsilon_n = -1/n^2$. Standard twofold anisotropy ($m_1 = m_2 = m$; $m_3 > m$) corresponds to $\alpha_2 = 0$ and $0 < \alpha_1 < 1$. Values of α_1 (or α) in the range $-\infty < \alpha_1 < 0$ are also physically meaningful, corresponding to the case $m_1 = m_2 = m$ and $m_3 < m$. However, the same effect can be obtained by setting $\alpha_1 = \alpha_2$, with α_1 and α_2 both positive. In that case, $m_2 = m_3$ and $m_1 < m_3$ (or m_2). The extreme case of $\alpha_1 = 1$ and $\alpha_2 = 0$ corresponds to a pure

TABLE I. Calculated energies of a harmonic oscillator for the lowest four states for $\alpha_1 = \frac{8}{9}$ and $\alpha_2 = 0$ as a function of the highest angular momentum l_{\max} and the number of components N .

N	1	2	3	4	5	6	7	Exact
l_{\max}	0	2	4	6	8	10	12	
1s	2.5157	2.3590	2.3364	2.3337	2.3334			2.3333
2s	5.8788	4.2234	3.7914	3.6903	3.6703	3.6673	3.6668	3.6667
l_{\max}	1	3	5	7	9	11		
$2p_z$	3.4156	3.0742	3.0114	3.0017	3.0002	3.0001		3.0000
$2p_x$	4.5338	4.3684	4.3388	4.3340	4.3334			4.3333

two-dimensional symmetry with exact solutions, the ground state having an energy of $\epsilon = -4$. The other extreme case $\alpha_1 = \alpha_2 = 1$ reduces to a one-dimensional problem with the ground-state energy collapsing to ($E \rightarrow -\infty$).

To test the method, we have applied it to the anisotropic harmonic oscillator, with a potential analogous to that of Eq. (8):

$$V(\mathbf{r}) = (1 + \alpha_2)x^2 + (1 - \alpha_2)y^2 + (1 - \alpha_1)z^2. \quad (10)$$

The eigenenergies have the analytic form

$$E = \sqrt{1 + \alpha_2}(2n_1 + 1) + \sqrt{1 - \alpha_2}(2n_2 + 1) + \sqrt{1 - \alpha_1}(2n_3 + 1),$$

where n_1, n_2, n_3 are nonnegative integers. Table I shows the systematic convergence of the lowest four energies for $\alpha_2 = 0$ and $\alpha = \alpha_1 = \frac{8}{9}$. The calculated energies are listed together with the number of components (N) and the maximum angular momentum (l_{\max}) used in a calculation. States containing only even l and $m = 0$ components are labeled s states; those with odd l and $m = 0$, p_z states and those with odd l and $m = 1$, p_x states. For the energy to converge to within $\delta\epsilon = 10^{-4}$, the 1s, $2p_x$, $2p_z$, and 2s states require $N = 5, 5, 6$, and 7 ($l_{\max} = 8, 9, 11$, and 12) components, respectively. While four components are sufficient to allow the ground states of each symmetry (1s, $2p_x$, and $2p_z$) to converge to within $\delta\epsilon = 0.001$, the 2s excited state requires six components. One important result is the uniform convergence of these calculations: Each increase in the number of components produces an energy which is both lower and closer to the exact value than its predecessors.

It is worth noting that $\alpha = \frac{8}{9}$ represents a strong anisotropy. As one would expect, significantly lower values of α require much smaller numbers of components N . For threefold anisotropy, the l_{\max} required for convergence is roughly that needed in a twofold anisotropic problem with an α equal to the larger of two threefold anisotropy parameters α_1 and α_2 . Of course, the total number of components N needed for a given l_{\max} is now considerably larger. s states would now include components with l even, m even and positive; p_z states l odd, m even and positive; p_x states l odd, m odd and positive; and p_y states l odd, m odd and negative.

For the anisotropic-mass Coulomb problem, the speed of convergence is close to that for the harmonic oscillator. Table II shows the convergence of the $2p_z$ and $3p_z$ energies for the twofold anisotropies $\alpha = 0.488$ and 0.973. For $\alpha = 0.488$, both states converge to within $\delta\epsilon = 0.001$ or 0.002 with only two components ($l_{\max} = 3$) included. For the strong anisotropy $\alpha = 0.973$, five components are sufficient to produce an accuracy of $\delta\epsilon \leq 0.002$. Table III compares energies we obtained with $N = 5$ with the best previous results.⁶ The values are nearly equal in most cases, while there are a few cases with significant differences. In these cases, our calculation yields the lower energies.

The tests and comparisons described above allow us to proceed to the Coulomb potential with threefold mass anisotropy with considerable confidence. As was mentioned before, the two parameters in the range $0 \leq \alpha_1 \leq 1$ and $0 \leq \alpha_2 \leq 1$ cover the entire domain of threefold anisotropy. By analogy to the previously discussed convergence behavior, the energies we present are accurate to within about $\delta\epsilon = 0.002$. Table IV gives our calculated energies

TABLE II. Convergence of the $2p_z$ and $3p_z$ energies ϵ from Eq. (8) in the twofold anisotropic Coulomb potential for $\alpha = 0.488$ and 0.973.

N	1	2	3	4	5	6	7	8
l_{\max}	1	3	5	7	9	11	13	15
$\alpha = 0.488$								
$2p_z$	-0.36	-0.366	-0.3665	-0.3665				
$3p_z$	-0.16	-0.1652	-0.1653	-0.1653				
$\alpha = 0.973$								
$2p_z$	-0.94	-1.16	-1.225	-1.244	-1.2503	-1.2520	-1.2528	-1.2528
$3p_z$	-0.42	-0.594	-0.671	-0.705	-0.7145	-0.7150	-0.7155	-0.7156

TABLE III. Calculated energies ϵ in a Coulomb potential with twofold mass anisotropy, comparing present values with those of Faulkner (Ref. 6).

	1s	2s	2p _z	3p _z	2p _x	3p _x
			$\alpha=0.784$			
Present	-1.553	-0.438	-0.566	-0.269	-0.320	-0.155
Faulkner	-1.553	-0.437	-0.565	-0.269	-0.320	-0.155
			$\alpha=0.936$			
Present	-2.012	-0.704	-0.965	-0.497	-0.361	-0.210
Faulkner	-2.01	-0.695	-0.933	-0.496	-0.361	-0.210
			$\alpha=0.973$			
Present	-2.321	-0.952	-1.250	-0.715	-0.384	
Faulkner	-2.31	-0.928	-1.24	-0.715	-0.384	

for combined values of α_1 and α_2 in which the two parameters are allowed to range independently over 0.0, 0.2, 0.4, 0.6, and 0.8. Figure 1 plots the energies of the 1s state as a function of α_1 for four values of α_2 . In each of Figs. 2(a)–2(e), we plot the calculated energies of the 2s, 2p_x, 2p_y, and 2p_z as a function of α_1 for a given value of α_2 . The magnitudes of the energies increase more than quadratically as a function of both parameters for all but the 2p_x state. 2p_z is the state most sensitive to variation in α_1 , while 2p_x is the least. It is the 2p_y state which is

most sensitive to variation in α_2 . The 2p_y and 2p_z states cross when $\alpha_1 = \alpha_2$, as can be predicted from the form of the potential. These results should be useful for qualitative analysis of the donor levels in semiconductors with three distinct masses.

IV. SUMMARY AND DISCUSSION

In this paper, we presented a general method for solving the eigenvalue problem of a Hamiltonian with a non-

TABLE IV. Calculated energies ϵ of a particle with a threefold anisotropic mass in a Coulomb potential.

	α_2	α_1	0.0	0.2	0.4	0.6	0.8
1s		0.0	-1.000	-1.010	-1.042	-1.109	-1.252
		0.2	-1.076	-1.086	-1.122	-1.195	-1.354
		0.4	-1.176	-1.188	-1.228	-1.311	-1.492
		0.6	-1.322	-1.337	-1.384	-1.482	-1.695
		0.8	-1.581	-1.600	-1.661	-1.786	-2.063
2s		0.0	-0.250	-0.254	-0.267	-0.297	-0.377
		0.2	-0.269	-0.273	-0.287	-0.318	-0.401
		0.4	-0.298	-0.302	-0.316	-0.347	-0.434
		0.6	-0.345	-0.349	-0.363	-0.396	-0.485
		0.8	-0.452	-0.456	-0.472	-0.505	-0.595
2p _z		0.0	-0.250	-0.252	-0.255	-0.264	-0.281
		0.2	-0.284	-0.287	-0.292	-0.302	-0.322
		0.4	-0.336	-0.338	-0.344	-0.357	-0.383
		0.6	-0.417	-0.420	-0.429	-0.446	-0.482
		0.8	-0.587	-0.591	-0.605	-0.634	-0.693
2p _x		0.0	-0.250	-0.233	-0.222	-0.216	-0.216
		0.2	-0.261	-0.243	-0.232	-0.225	-0.225
		0.4	-0.275	-0.256	-0.243	-0.236	-0.237
		0.6	-0.294	-0.273	-0.259	-0.249	-0.252
		0.8	-0.323	-0.299	-0.283	-0.274	-0.276
2p _y		0.0	-0.250	-0.274	-0.311	-0.371	-0.495
		0.2	-0.261	-0.287	-0.325	-0.390	-0.523
		0.4	-0.275	-0.303	-0.344	-0.414	-0.560
		0.6	-0.294	-0.324	-0.370	-0.446	-0.610
		0.8	-0.323	-0.357	-0.410	-0.499	-0.693

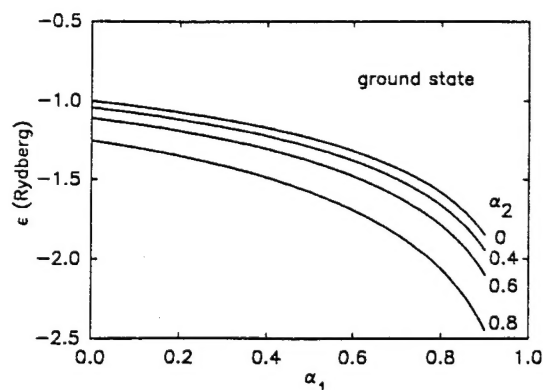


FIG. 1. Ground-state energy of the Coulomb potential with threefold mass anisotropy as a function of α_1 for $\alpha_2 = 0.0, 0.4, 0.6$, and 0.8 .

spherical potential and/or masses and obtained results for a hydrogenic potential with threefold mass anisotropy. Each eigenfunction is represented as a linear combination of basis functions, each of which is itself expanded in spherical harmonics. The nonspherical potential couples the radial functions for a given basis function, generating a set of coupled radial differential equations which we solve numerically. The eigenenergies are obtained by varying the energy to match the value and radial first derivative of the inner and outer parts of the wave function at a chosen radius.

We test the applicability of this method and its speed of convergence against the analytical energies of the anisotropic harmonic oscillator and against previous results obtained for the Coulomb potential with twofold mass anisotropy. The results show that the method is fully

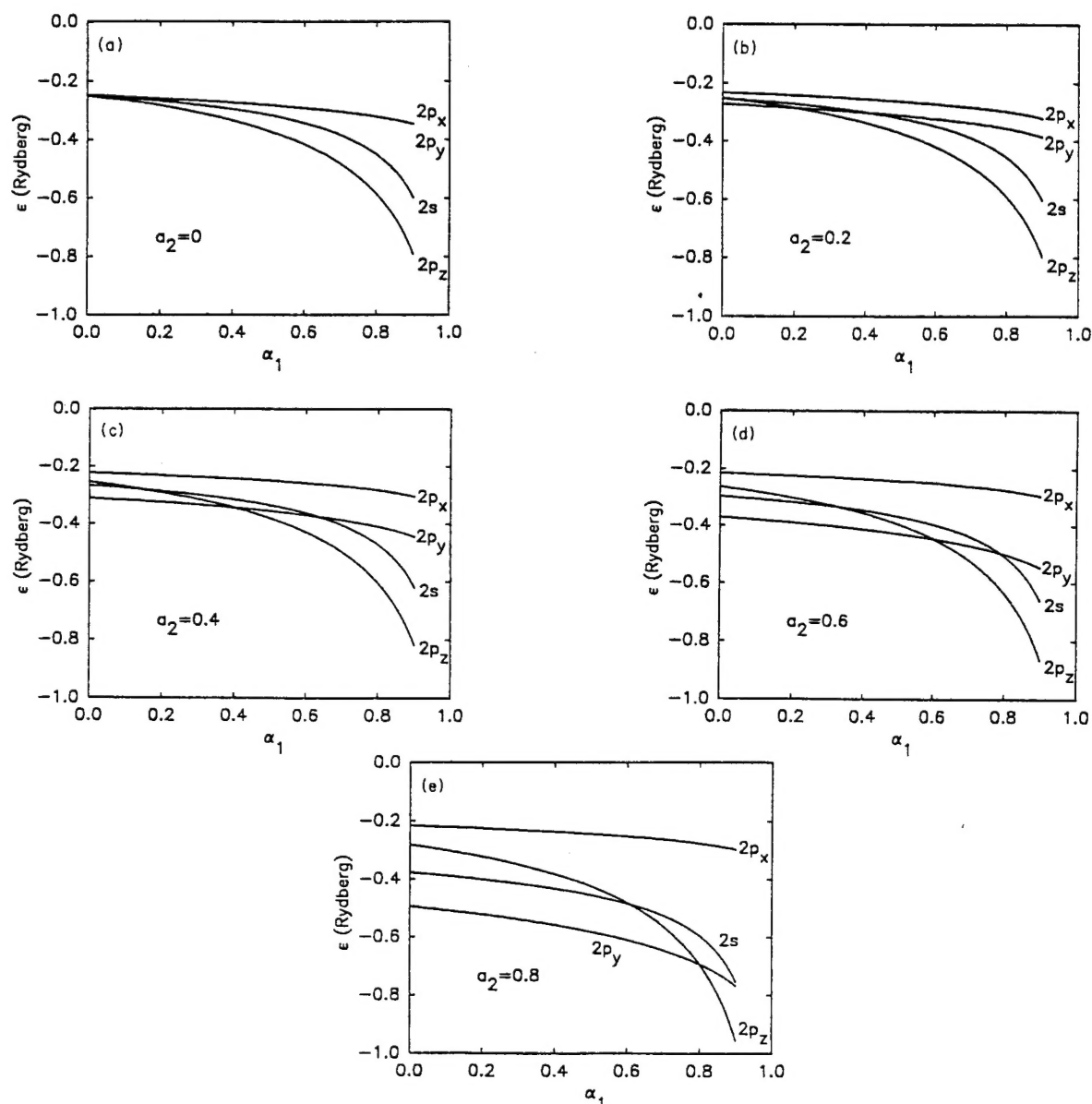


FIG. 2. Lower excited-state energies of the Coulomb potential with threefold mass anisotropy as a function of α_1 . (a), (b), (c), (d), and (e) are for $\alpha_2 = 0.0, 0.2, 0.4, 0.6$, and 0.8 , respectively.

variational, the energy obtained always being higher than the exact value for both the ground and excited states. Our results also show monotonic convergence, with the energy being lowered as the number of independent basis functions increases, until the exact energy is reached. On the basis of these tests, we are able to find accurate energies for the hydrogenic potential with threefold mass anisotropy over a wide range of the anisotropy parameters.

Although the hydrogenic potential with effective masses provides a useful qualitative model for describing the energy levels of shallow donors and acceptors in semiconductors, the model and the present method are not suitable for detailed study of these impurity states. When the impurity state is expanded in terms of the band states around the band edge, the effective potential thus generated is a product of the bare impurity potential and a periodic function derived from the band-edge Bloch functions. Even when the bare impurity potential has a single minimum at the impurity site, the Bloch functions will produce many subsidiary minima in the full effective potential. This would invalidate the present method, which is based on an expansion around a single center. Whether the present method can be modified to deal with an extended impurity potential with multiple minima is not clear at the present time.

One of the most important aspects of this work is in the construction of the basis functions, which also plays an important part in the full-potential (FP) KKR theory.^{1,5} While most of the work done on this theory employs integral equations with the free-electron Green function as the kernel, the present method solves a set of coupled differential equations to obtain the basis functions. As was mentioned earlier, if one starts with a set of independent basis functions at a small radius and integrates outward, the basis functions tend to lose their linear independence by the time one reaches the matching radius. This occurs whenever there are three or more components in a basis function, for weak as well as strong

anisotropy, and for the harmonic oscillator as well as for the Coulomb potential. Since the harmonic-oscillator potential is not singular, this behavior is not, or is not solely, due to the potential singularity described by Butler, Gonis, and Zhang.¹

We speculate that the problem is caused by the fact that the spherical component is not dominant at small r . This allows strong mixing of all the components of a basis function at small radii, which in turn leads to loss of independence. This problem has not arisen in previous calculations testing FP KKR theory, since most of these calculations used model potentials which were either spherical or dominated by the spherical term at small radii. Other tests⁵ have used a small set of basis functions in potentials with high symmetries, such as cubic harmonics, which did not couple more than two radial components—a situation in which the loss of linear independence also does not occur in our method. We expect, however, that the FP KKR method will encounter potentials more like the potentials examined here when it is applied to open structures, such as most semiconductors. One good example is the potential for interstitial cells. This problem deserves further investigation.

As we have shown by our present calculations, however, the problem of linear independence of the basis functions can be circumvented. This is done by orthogonalizing the basis functions at several values of r in the small- r region during integration outward. We believe our method will facilitate the execution of FP KKR theory for semiconductors and other open structures.

ACKNOWLEDGMENTS

We thank P. Srichaikul and Dr. J. W. Choyke for useful discussions on SiC. This work was supported in part by NASA Grant No. NAGW-1192 and DARPA Contract No. MDA972-92-C-0053.

¹W. H. Butler, A. Gonis, and X.-G. Zhang, *Phys. Rev. B* **45**, 11 527 (1992), and references therein.

²P. Srichaikul and A.-B. Chen (unpublished).

³J. Korrington, *Physica* **13**, 392 (1947).

⁴W. Kohn and N. Rostoker, *Phys. Rev.* **94**, 1111 (1954).

⁵C.-Y. Yeh, A.-B. Chen, D. M. Nicholson, and W. H. Butler,

Phys. Rev. B **42**, 9336 (1990).

⁶R. A. Faulkner, *Phys. Rev.* **184**, 713 (1969).

⁷J. Pollman, *Solid State Commun.* **19**, 361 (1976).

⁸D. Wintgen, H. Marxer, and J. S. Briggs, *J. Phys. A* **20**, L965 (1987).

⁹J.-B. Xia, *Phys. Rev. B* **39**, 5386 (1989).